Extracellular Signalling and Stem Cell Self-Renewal

Thesis

Extracellular Signalling and Stem Cell Self-Renewal

Sissy E. Wamaitha, BSc (Hons)

The Francis Crick Institute
(Formerly MRC National Institute for Medical Research)

Thesis Submitted for the Degree of
Doctor of Philosophy

The Open University July 2017

Declaration

The work contained herein is my own, except where explicitly stated otherwise, and represents an original report of my research. This work has not been previously submitted for a degree or other qualification at this or another institution. Portions of this work were presented and published in research article form in the following jointly-authored publications:

Wamaitha SE; Del Valle I; Cho LTY; Wei Y; Fogarty NME; Blakeley P; Sherwood RI; Ji H; Niakan KK (2015). Gata6 potently initiates reprograming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. *Genes & Development*, 29, 1239-1255

Blakeley P, Fogarty NME, Del Valle I, Wamaitha SE, Hu TX, Elder K, Snell P, Christie L, Robson P, Niakan KK (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development*, 142(18), 3151-3165

For the sake of completeness in some areas of the subject matter, data generated in collaboration with or by other individuals has been included in this text. Any such contribution from my colleagues, co-authors or collaborators, or any information derived from other sources, is explicitly indicated and referenced within the body of the text, or in figure and/or table legends as appropriate.

Human embryos surplus to IVF were donated to the research project by informed consent under the UK Human Fertilisation and Embryology Authority Licence number R0162. Ethical approval was sought and granted (Research Ethics Committee (REC) reference:04/Q0108/99).

Acknowledgements

Over the course of my PhD, I have had the privilege of working alongside dedicated and knowledgeable researchers, to whom I owe an immense debt of gratitude. First and foremost, I wish to thank my supervisor Kathy Niakan, for her invaluable expertise and her constant support and guidance. Thank you for taking a chance on me those many years ago. Sincere thanks to Nacho del Valle, Norah Fogarty and Paul Blakeley, not only for their experimental contributions and intellectual discussions, but also for making long days in the lab pass quicker with the joy of their company. I am eternally grateful to my second supervisor, James Turner, for his words of advice and encouragement, and for being a true friend. I would also like to thank my thesis committee members, Francois Guillemot and Robin Lovell-Badge, for their insightful critiques and helpful advice during the course of my PhD, and the Medical Research Council for providing the funding for me to carry out my research.

I gratefully acknowledge the Guillemot, Hadjantonakis, Rossant, Smith and Ultanir labs for the mouse cell lines they supplied for use in this project, and Russell Ernst in the Chin lab at the MRC-LMB, for his help generating the Erk2-Venus reporter lines. I am indebted to the confocal microscopy, high-throughput sequencing, and Human Embryo and Stem Cell core facilities, both at the MRC National Institute for Medical Research, and now at The Francis Crick Institute – and especially to Donald Bell, Abdul Sesay, Leena Bhaw-Rosun, Nicolle Morey, Marta Miret, Lucia Goyeneche and Daria Beldzik for their assistance. I would also like to thank Christophe Galichet, for suffering my questions. I am immensely grateful to our IVF clinic partners and the anonymous donors, without whose support this research would not be possible.

Finally, I would not have been able to complete this endeavour without the friendship and support of a number of people. My thanks go to: Brook Cooper and Mike Doran, for keeping the legacy of Boam alive, and introducing me to so many wonderful people, too many to list definitively here. To Darryl Hayward, Joe Wright and Ola Lubojemska, and Bad-Idea-Bears Lilly

Hunt and Helen O'Neill, for much-needed hilarity. To Daniel Snell, for several excellent conversations, scientific and otherwise. To Donald 'JFDI' Bell, for your sunny disposition, and occasionally being there when I need you. To Luuk Baltussen and Bill Wray, for some high-quality Halloweens. To Matthew Williams and Guy Hallifax, for all your help. To NIMDRAM, for the revelation that amateur dramatics can sometimes be more demanding than a PhD. To Sophia Davidson, Teresa McCabe and Wiebke Nahrendorf, for CapeGang™, crafternoons and shouldering the burdens of inappropriate footwear.

Somewhat further afield, I am grateful to Alex Granata and Lucy Major (née Low) for the pleasure of their company. To Will Bernard, for the inappropriate laughs. To Fiona Docherty, for our adventures, past and future, and to Josh McTigue, for our band-in-the-making. To Matthias Hofer, for a hundred coffees and endless conversations. To Stanley Strawbridge, for all the things we don't talk about, and all the things we do. To Nadia Tyler-Rubinstein, for every immersive experience and the joys of Accidental Tuesdays. To Liam Simmonds and Ruby Newton, for all the time spent patiently waiting for me to get out of cell culture or collect my timepoints, when we should have been having fun. And to Ashwin Lee, Huzaira Khan (née Bashir), Priscilla Kjizi José, Brian Waihiga, Simon Mumbo and Jonathan Ochako, for still being there.

Finally, to my family for their love and support, and most of all, their understanding.

Abstract (299 words)

In preimplantation mouse embryos, signalling and gene regulatory networks cooperate to determine lineage segregation, and modulating signalling *in vitro* allows for stem cell populations to be established from these lineages. Fibroblast growth factor (FGF) signalling triggers the differentiation of primitive endoderm (PrE) cells fated to contribute to the yolk sac, while cells unreceptive to FGF form the epiblast (Epi) that subsequently contributes to the embryo proper. *In vitro*, FGF signalling is required for preimplantation Epi-derived mouse ES cells to exit self-renewal. Conversely, in human ES cells and postimplantation Epi-derived mouse epiblast stem cells, FGF signalling is instead required for pluripotency maintenance. It remains unclear how these divergent outcomes arise, especially as these cells rely on a similar core pluripotency gene network.

This study demonstrates that ectopic expression of the PrE transcription factor Gata6 destabilises mouse ES cell pluripotency *in vitro* and upregulates PrE-associated genes independently of FGF signalling. As previous studies show that PrE specification is compromised in *Fgf4*$^{-/-}$ embryos, despite initiation of Gata6, this suggests FGF signalling and Gata6 cooperatively drive PrE specification *in vivo*. Characterising Gata6 function determines that it directly binds to both up- and downregulated gene targets and potently initiates reprogramming in multiple cell types, including human ES cells, suggesting it may also antagonise pluripotency *in vivo*.

Surprisingly, FGF stimulation negatively affects establishment of the pluripotent human Epi. Characterising alternative signalling pathways in the human embryo finds that modulating IGF signalling promotes proliferation of the human ICM, and similar to human ES cells, intact TGFβ/Nodal signalling is required for pluripotent gene expression in the Epi. Consequently, as signalling requirements in the human Epi appear somewhat distinct from both the mouse Epi and existing human ES cells, modulating embryo-specific signalling pathways may permit derivation of human ES cells that more accurately reflect the pluripotent Epi compartment.

Table of Contents

Table of Illustrations

Introduction

Results Chapter 1

Results Chapter 2

Table of Tables

Abbreviations

BMP – Bone morphogenetic protein

ChIP-seq – Chromatin immunoprecipitation sequencing

cXEN cell – Converted XEN cell (e.g. by growth factor modulation)

DE – Definitive endoderm

E2.5 – Embryonic day 2.5 (2.5 days post fertilisation)

EB – Embryoid body

EC – Embryonal carcinoma cell

Epi – Epiblast

ES cell – Embryonic stem cell

ExEn – Extraembryonic endoderm

eXEN cell – Embryo-derived XEN cell

FGF – Fibroblast growth factor

GEMS – Genetic and Embryo Manipulation Services (Crick Facility)

ICM – Inner cell mass

IHC – Immunohistochemistry

iPS cell – Induced pluripotent stem cell

iXEN – Induced XEN cell

JAK/STAT – Janus Kinase / Signal Transducer and Activator of Transcription

KOSR – Knockout serum replacement

LIF – Leukaemia inhibitory factor

MAPK – Mitogen activated protein kinase

MEF – Mouse embryonic fibroblast (supportive cell layer)

MEF-CM – MEF conditioned media

MINS – Modular Interactive Nuclear Segmentation (Matlab-based software tool)

OE - Overexpression

PrE – Primitive endoderm

RNA-seq – Ribonucleic acid sequencing

RPKM – Reads per kilobase of exon model per million mapped reads

qRT-PCR – Quantitative reverse transcriptase polymerase chain reaction

SB – SB-431542, TGFβ signalling inhibitor

shRNA – Short hairpin ribonucleic acid

STO – SIM (Sandos Inbred Mice) Thioguanine/Ouabain-resistant mouse fibroblast cell line

TE – Trophectoderm

TF – Transcription factor

TS cell – Trophectoderm stem cell

WB – Western blot

XEN cell – Extraembryonic endoderm stem cell

Introduction

1. Lineage specification in the preimplantation embryo

Preimplantation mouse and human embryos exhibit remarkably similar morphology, and development at this stage involves a series of cell fate choices as fertilised 1-cell zygotes undergo a number of mitotic cell divisions; three rounds of cleavage to form an 8-cell embryo, followed by compaction and cavitation to form a blastocyst (reviewed in (Kojima et al. 2014b)). However, there are a number of significant distinctions in human development (De Paepe et al. 2014), such as a longer temporal period for blastocyst formation, and different timing of major zygotic/embryo genome activation, which occurs between the 2- and 4-cell stages in the mouse (Hamatani T. 2004) but between the 4- and 8-cell stages in human embryos (Niakan et al. 2010). During early embryonic development, cells become gradually restricted in their differentiation potential, with cell identity determined by position, polarisation, asymmetric cell division and lineage-specific gene expression (Schrode et al. 2014; Chazaud and Yamanaka 2016). A timeline of preimplantation development in mouse and human ES cells is shown in Figure A.

The first lineage specification event in the mouse is the segregation of the inner cell mass (ICM) and the trophectoderm (TE), with the latter contributing solely to the foetal portion of the placenta. This is followed by the specification of the cells within the ICM to either pluripotent epiblast (Epi) progenitor cells, which form the embryo proper, or to primitive endoderm (PrE) cells, which contribute predominantly to the yolk sac (Gardner et al. 1973; Gardner and Rossant 1979; Gardner 1985). These contributions were assessed by analysis of mouse postimplantation phenotypes, which for ethical reasons cannot be carried out in the human, but it is likely that these assignations hold true in a human context. Each lineage is characterised by a network of supporting genes, mainly transcription factors, whose role was largely identified through mouse mutant phenotypes affecting lineage emergence or maintenance. The present study will focus mainly on the emergence of the Epi and PrE lineages within the ICM, rather than on the TE.

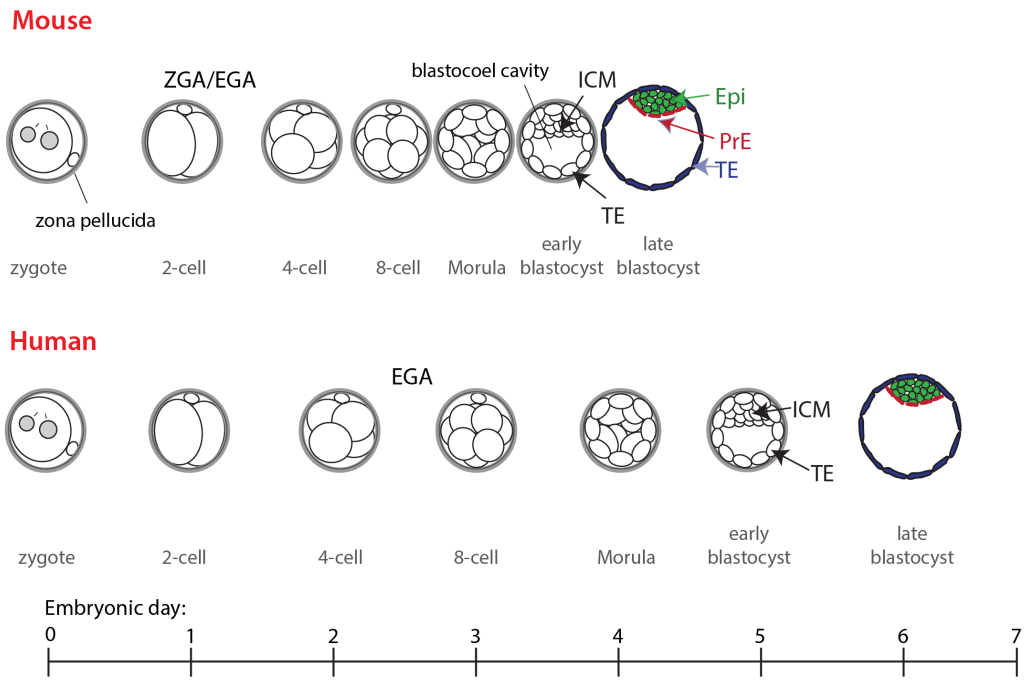


Figure A: A comparison of mouse and human preimplantation development. Early embryonic development in both mice and humans is characterised by a series of cell divisions resulting in the formation of a blastocyst comprising a discernible inner cell mass (ICM) and trophectoderm (TE). This occurs between embryonic day 3 and 4 in mice, and day 5 and 6 in humans. The ICM is comprised of a pluripotent epiblast (Epi) that will give rise to all 3 germ layers of the foetus, and a primitive endoderm (PrE) layer that gives rise to the extraembryonic endoderm (ExEn) cells that will form the yolk sac. In addition to differences in the timing of blastocyst formation, mouse and human embryos differ significantly in the timing of major zygotic/embryo gene activation (ZGA/EGA) and compaction, with human embryos being comparatively delayed. Figure adapted from Niakan et al., 2010, *Development*, doi: 10.1242/dev.060426. Reproduced with permission of Company of Biologists, in the format Republish in a thesis/dissertation via Copyright Clearance Center, confirmation Number: 11627816.

In the mouse Epi this network is centred on Oct4 (*Pou5f1*), Sox2 and Nanog, while the PrE instead requires an extraembryonic endoderm (ExEn) program, including transcription factors Gata6 and Sox17. Both *Pou5f1*$^{-/-}$ and *Nanog*$^{-/-}$ embryos fail to form a pluripotent epiblast (Chambers et al. ; Nichols et al. 1998; Mitsui et al. 2003), and all cells in the *Nanog*$^{-/-}$ ICM instead express Gata6 (Frankenberg et al. 2011). *Sox2*-null embryos appear normal until shortly after implantation, however this is thought to be due to persistent maternal Sox2 protein (Avilion et al. 2003). Both *Pou5f1*-null or *Sox2*-null mutant ICM outgrowths only give rise to trophectoderm cells (Nichols et al. 1998; Avilion et al. 2003). Although previously thought to have a postimplantation ExEn phenotype (Morrisey et al. 1998; Koutsourakis et al. 1999), recent work has shown that *Gata6*$^{-/-}$ embryos fail to form the PrE, and Oct4, Nanog, and Sox2 are instead expressed across the ICM (Bessonnard et al. 2014; Schrode et al. 2014). Fewer PrE cells are observed in heterozygous Gata6$^{+/-}$ embryos, indicating a threshold of Gata6 expression is required for PrE specification. Although *Sox17*$^{-/-}$ embryos form a PrE layer, PrE cells are progressively lost if implantation is delayed in these embryos (Artus et al. 2011), suggesting Sox17 is required for PrE maintenance.

These distinct gene expression patterns were confirmed by genome-wide expression analysis (Kurimoto et al. 2006; Guo et al. 2010) and provide a useful means of tracking lineage specification during preimplantation development, especially in the mouse. Oct4 is expressed in all cells within the ICM, and only becomes restricted to the Epi in the late blastocyst (Grabarek et al. 2012). Nanog and Gata6 are the earliest markers of the presumptive Epi and PrE, and are initially co-expressed before resolving into a mosaic salt-and-pepper pattern in the ICM (Plusa et al. 2008). Experiments using *Pdgfra*$^{H2B\text{-}GFP}$ reporter mice to mark presumptive PrE cells showed that ICM cells initially randomly segregate, with Epi or PrE commitment correlated with Nanog or Gata4 upregulation respectively, and cell migration and subsequent apoptosis ensuring that cells are correctly located (Plusa et al. 2008).

The mechanisms of lineage specification in the human ICM are much less well characterized, although some gene expression patterns are shared with the mouse (De Paepe et al. 2014). OCT4 is initially expressed in all cells before being restricted to the human Epi in the late blastocyst (Niakan and Eggan 2013), along with NANOG and SOX2 (Cauffman et al. 2009). Furthermore, SOX17 is localized to the PrE, where it overlaps with GATA6 (Niakan and Eggan 2013). However, although the TE lineage specifier Cdx2 is expressed from the morula stage in the mouse, CDX2 was only detected after cavitation in the human blastocyst, suggesting lineage specification mechanisms may not necessarily be conserved between the two species. In contrast, cynomolgus monkey embryos exhibit remarkably similar OCT4 and CDX2 expression dynamics to those in the human embryo (Nakamura et al. 2016), which may indicate the existence of a primate lineage specification mechanism that is distinct from rodents. Intriguingly, single cell transcriptomic analysis of cynomolgus monkey (Nakamura et al. 2016) and human embryos (Petropoulos et al. 2016) seemingly identifies the TE, Epi and PE emerging concurrently at the blastocyst stage, rather than in two sequential steps as in the mouse. Several Epi- and TE-associated genes are initially co-expressed in the human embryo before being restricted to their specific lineages (Petropoulos et al. 2016). This may reflect greater plasticity during early primate development, and both inner and outer cells of blastocysts disaggregated at embryonic day 5 (E5) are capable of forming a blastocyst with both an ICM and TE (De Paepe et al. 2013). Further analysis of protein expression patterns during development will be crucial to support these conclusions, especially with regards to novel lineage markers that are highly enriched in the human embryo (such as ARGFX and KLF17 in the Epi, FOXA2 in the PE, and GATA2 and GATA3 in the TE, (Blakeley et al. 2015; Petropoulos et al. 2016)).

2. Signalling in the early embryo – defining the Epi and PrE

Extracellular signalling has been linked to Epi and PrE specification in the mouse, but it is unclear to what extent this is conserved in the human. In the mouse embryo, fibroblast growth factor (FGF) signalling, and consequent activation of the mitogen-activated protein kinase

(MAPK) pathway, is required to facilitate the segregation of Epi and PrE lineages (Lanner and Rossant 2010). A complementary receptor-ligand relationship has been proposed, where the FGF receptor Fgfr2 is thought to be more highly enriched in PrE cells, while EPI cells secrete an FGF ligand, Fgf4, (Rappolee et al. 1994; Goldin and Papaioannou 2003; Guo et al. 2010). In addition, mutations in genes coding for FGF signalling pathway components result in defects in PrE formation, phenocopying the effects seen with mutations in various lineage specifiers. *Fgfr*$^{-/-}$ and *Fgf4*$^{-/-}$ mutant embryos fail to form a PrE layer (Feldman et al. 1995; Arman et al. 1998), as do *Grb2*$^{-/-}$ mutant embryos, where loss of the Grb2 adaptor protein affects FGF signal transduction (Cheng et al. 1998).

In addition to morphological defects, disrupting FGF signalling also affects gene expression patterns within the ICM. Gata6 expression is lost in the *Grb2*$^{-/-}$ E3.5 ICM, where all cells become Nanog-expressing epiblast progenitors (Chazaud et al. 2006). Inhibiting FGF receptors or the MAPK Mek also increases the proportion of Nanog-expressing cells within the ICM (Nichols et al. 2009; Yamanaka et al. 2010), suggesting that FGF signalling primarily functions via the Raf-Mek-Erk pathway. Both Mek1 and Erk2 mutants have an embryonic lethal phenotype in mice (Giroux et al. 1999; Hatano et al. 2003) suggesting a key and distinct role for these kinases during early development. Conversely, treatment with FGF ligands alone or coupled with the FGF receptor binding facilitator heparin resulted in downregulation of Nanog expression and conversion to Gata6-positive PrE progenitors (Yamanaka et al. 2010).

However, it is unclear how FGF signalling interacts with either the pluripotency or the ExEn gene regulatory networks in the mouse embryo. Recent studies show that Gata6 is initiated in *Fgf4*$^{-/-}$ mutant embryos at E2.0, but subsequent activation of PrE factors Gata4 and Sox17 is compromised (Kang et al. 2013). Although it was previously reported that Gata6 is not expressed in *Grb2*$^{-/-}$ embryos at E3.5 (Chazaud et al. 2006), by this time Gata6 expression has diminished in *Fgf4*$^{-/-}$ embryos, making it likely that Gata6 is also initiated but not maintained in *Grb2*$^{-/-}$

embryos. Although ectopic expression of Gata6 in *Grb2*$^{-/-}$ embryos can rescue the PrE differentiation defect (Wang et al. 2011), exogenous FGF signalling is not sufficient to drive Gata4 and Sox17 expression in *Gata6*$^{-/-}$ embryos (Schrode et al. 2014). Furthermore, modulating FGF signalling at E3.5, after Gata6 or Nanog were expressed, did not affect lineage specification. This seemingly suggests that Gata6 lies downstream of FGF signalling to drive a PrE fate.

The signalling pathways responsible for lineage specification in the human ICM are less well characterised, despite conservation of lineage-specific genes between mouse and human embryos. Curiously, inhibiting FGF signalling in human embryos (Kuijk et al. 2012; Roode et al. 2012) has no effect on either PrE or Epi formation in the human blastocyst, and gene expression patterns are unchanged. However, Nodal- or Activin-driven TGFβ signalling has been shown to have a role in maintaining pluripotency *in vitro*, which will be discussed below, and inhibiting TGFβ/Nodal in the human embryo affects the expression of NANOG (Blakeley et al. 2015).

Overall, gene networks and signalling pathways appear to intersect in order to determine and maintain lineage fate in both the mouse and the human embryo. Consequently, appropriately modulating signalling pathways has enabled the establishment of cell lines *in vitro* that retain the characteristics of their embryonic cell type of origin.

3. Capturing pluripotency *in vitro*

A number of observations provided the foundations for the derivation and subsequent appreciation of embryonic stem cells. It was noted that scrotal tumours spontaneously arose in 129 strain mice (Stevens and Little 1954; Stevens 1967), and that these teratomas were comprised of a mixture of several different cell types. These masses could be transplanted to colonise other sites, and were thus designated as teratocarcinomas, with the hypothesis that the constituent embryonal carcinoma (EC) cells were a type of undifferentiated stem cell (Kleinsmith and Pierce 1964). Supporting this, single EC cells were shown to be sufficient to generate all the divergent

tissues observed in teratocarcinomas (Kleinsmith and Pierce 1964; Martin and Evans 1974), and were thus pluripotent. Cells isolated from the mouse ICM were shown to have a similar antigen expression pattern to mouse EC cells, suggesting they too may be pluripotent. Consequently, conditions used to culture mouse EC cells were applied to attempt to generate stable cell lines from mouse preimplantation embryos.

Deriving mouse embryonic stem (ES) cells

The first stable mouse ES cells were derived from 129 strain mice in two concurrent studies. Diapause-induced blastocysts were plated intact in microdrops of serum-containing medium and colonies reminiscent of EC cells picked and passaged to derive a stable cell line (Evans and Kaufman 1981). The resulting cells were karyotypically normal and contributed to teratocarcinomas when injected subcutaneously into recipient adult mice. In the alternative study, isolated ICMs from normal blastocysts were plated onto a STO feeder layer in serum-containing feeder-conditioned medium (Martin 1981). Again, EC-like colonies emerged and could be maintained, and gave rise to teratocarcinomas *in vivo*, and to multiple differentiated tissues in EB assays *in vitro*.

The term "embryonic stem cell" was coined to distinguish these cells from their EC counterparts and to reflect their embryonic origin (Martin 1981). Subsequent studies showed that mouse ES cells successfully contributed to chimaeras (Bradley et al. 1984), predominantly to foetal and extraembryonic mesoderm tissue, and with only a low frequency to the TE and PrE (Beddington and Robertson 1989). Mouse ES cells depend on a gene regulatory network surrounding the core transcription factors Oct4 (*Pou5f1*), Sox2, and Nanog (Boyer et al. 2005; Loh et al. 2006; Chen et al. 2008), similar to the Epi compartment from which they were initially derived. Additional transcription factors such as Esrrb and Klf4 are also regulated by and reinforce the core pluripotency factors (Ivanova et al. 2006; van den Berg et al. 2008; Zhang et al. 2008; Hall et al. 2009; Niwa et al. 2009; Festuccia et al. 2012). Mouse ES cells that lack *Pou5f1* or *Sox2* lose

expression of pluripotency genes and differentiate primarily into trophectoderm stem (TS) cells (Niwa et al. 2000; Avilion et al. 2003; Mitsui et al. 2003; Ivanova et al. 2006; Masui et al. 2007). Although, *Nanog* mutant mES cells have been established by gene deletion *in vitro* they can only be maintained if leukaemia inhibitory factor (LIF) signalling is intact (Chambers et al. 2007).

LIF signalling was the first pathway identified as required for pluripotency maintenance. Briefly described as the differentiation inhibitory activity (DIA) factor that was secreted by MEFs to maintain mouse ES cells (Smith et al. 1988; Williams et al. 1988; Stewart et al. 1992), exogenous LIF was shown to replace the requirement for a MEF supportive layer in ES cell maintenance and derivation (Nichols et al. 1990). LIF maintains pluripotency via activation of JAK/STAT signalling, specifically Stat3, which triggers downstream gene expression to block differentiation towards mesoderm and endoderm lineages (Niwa et al. 1998; Bourillot et al. 2009). Serum and LIF are still routinely used to maintain ES cells in culture. In addition, bone morphogenetic protein (BMP) signalling can substitute for serum in mouse ES cell maintenance, acting via Smad and Id proteins to cooperate with LIF to maintain pluripotency (Ying et al. 2003).

Consistent with the mouse embryo, inhibiting FGF/Erk signalling in mouse ES cells also maintains pluripotency (Burdon et al. 1999; Kunath et al. 2007; Stavridis et al. 2007). Self-renewing pluripotent ES cells can be derived from *Grb2*$^{-/-}$ mouse embryos, but are refractory to endoderm differentiation *in vitro*, which can be restored by introducing a Sos1-Grb2 fusion to transduce FGF signalling (Cheng et al. 1998; Hamazaki et al. 2006; Wang et al. 2011; Findlay et al. 2013). Furthermore, addition of FGF4 to *Pou5f1*$^{-/-}$ ICM outgrowths which do not express *Fgf4*, promotes proliferation of TS cells (Nichols et al. 1998), further indicating that FGF4 is dispensable for ES cell maintenance. Conversely, constitutive active Mek results in Nanog repression in mouse ES cells (Hamazaki et al. 2006), indicating that FGF seemingly exercises the same dynamics in mouse ES cells as it does during Epi and PrE segregation in the embryo.

Although ES cells had been successfully derived from 129-related mouse strains, several other strains remained refractory to ES cell derivation (Gardner and Brook 1997; Blair et al. 2011). Cell lines from some these alternative strains could only be generated by hormonally inducing diapause to mimic *in vivo* embryonic delay, and even then cells were only derived at very low efficiencies. Consequently, it seemed plausible that confounding factors in culture were disrupting the pluripotency network, perhaps via FGF/Erk signalling activation, which was demonstrably detrimental. As conditions for serum- and feeder-free defined cultures had now been established (Ying et al. 2003), but neither LIF nor BMP blocked Erk activation *in vitro*, it was speculated that perhaps directly modulating FGF/Erk signalling would enable improved propagation of mouse ES cells (Ying et al. 2008).

Culture in 2i and LIF – "naïve" mouse ES cells

The addition of Mek and Gsk3β inhibitors together with LIF (2i+LIF, originally 3i+LIF) allowed mouse ES cells to be propagated in defined medium in the absence of BMP, serum or a MEF supportive layer (Ying et al. 2008; Nichols et al. 2009). Interestingly, *Stat3*-null mouse ES cells can be derived and maintained in 3i conditions, though they differentiate rapidly when returned to conventional culture (Ying et al. 2008), indicating that the inhibitors bypass a requirement for Stat3 activation. Gsk3β inhibition prevents phosphorylation (and subsequent degradation) of β-Catenin, which although not essential to maintain pluripotency provides additional resistance to differentiation (Lyashenko et al. 2011; Wray et al. 2011). Mouse ES cells lacking the transcriptional repressor Tcf3 do not respond to Gsk3β inhibition and upregulate Esrrb and Nanog, suggesting that Gsk3β inhibition alleviates repression of Tcf3 gene targets essential for pluripotency differentiation (Wray et al. 2011; Martello et al. 2013).

Remarkably, 2i+LIF conditions overcame the barriers to derivation previously observed, allowing stable ES cells to be established from alternative mouse strains, and from rat embryos (Buehr et al. 2008; Nichols et al. 2009). ES cells in 2i+LIF are thought to represent a "ground

state" of pluripotency, and have been shown to more closely resemble cells in the mouse preimplantation ICM, as compared to cells in classical serum and LIF culture conditions which correlate with later preimplantation epiblast stages (Boroviak et al. 2014). Cells in 2i+LIF also exhibited less heterogeneity compared to their serum and LIF counterparts. Altogether, this strongly suggests that current ES cell culture conditions appropriately capture the mouse pluripotent state. However, it is unclear whether the same is true in a human context.

Deriving human ES cells

The first human ES-like cells were isolated from human embryos that were plated and allowed to hatch onto a human oviduct epithelial feeder layer, in the presence of human serum and LIF (Bongso et al. 1994). After hatching, the ICM clumps that attached to the monolayer were disaggregated and sub-cultured in an attempt to maintain a stable self-renewing line. Although cells retained stem cell-like morphology, normal karyotype and stained positive for alkaline phosphatase, they could not be maintained for more than two passages (Bongso et al. 1994).

However, similar experiments performed in non-human primates were more successful. Immunosurgery was carried out on rhesus monkey or marmoset blastocysts to remove the trophectoderm, allowing the intact ICM to be plated onto a MEF layer and cultured in serum and human LIF (Thomson et al. 1995; Thomson et al. 1996). The resulting cell lines stained positive for alkaline phosphatase and expressed cell-surface markers characteristic of human EC cells such as SSEA-3, SSEA-4, TRA-1-60 and TRA-1-81, which were lost following differentiation. Furthermore, cells injected into immunodeficient SCID mice (Thomson et al. 1995) or aggregated into embryoid bodies (Thomson et al. 1996) differentiated into all 3 embryonic germ layers, indicating that these cells retained the pluripotent properties of the ICM. These primate ES cell lines could be maintained for multiple passages, and remained undifferentiated in the presence of MEF layers, but differentiated or died in their absence despite

the presence of human LIF. Therefore, although human LIF was added during derivation, it was dispensed with for routine culture.

The work in non-human primates refined the techniques that enabled successful derivation of ES cells from human embryos. Human ICMs isolated in a similar fashion and plated onto a MEF layer in the presence of serum gave rise to the first stable ES cell lines (H1, H7, H9, H13, H14,) that could be propagated for multiple passages and had a normal karyotype (Thomson et al. 1998). These cells also expressed alkaline phosphatase and pluripotency cell-surface markers, and formed teratomas when injected into SCID mice. Similar results were observed in a subsequent study (Reubinoff et al. 2000), which also demonstrated that although LIF was initially used during cell line establishment, it had no effect on the growth or differentiation of established cultures. Consequently, although human ES cells expressed OCT4 (Reubinoff et al. 2000) and thus likely relied on a similar pluripotency gene network to mouse ES cells, the signalling pathways involved were possibly distinct. A number of human ES cell lines have subsequently been derived (Mitalipova et al. 2003; Cowan et al. 2004; Suemori et al. 2006; Aflatoonian et al. 2010), though the H1 and H9 lines are predominantly used in human ES cell studies (Loser et al. 2010; Kobold et al. 2015).

Culture in defined conditions

Although culture conditions using serum and MEF layers had enabled successful human ES cell derivation, this again introduced potential issues with batch-to-batch variability and exposure to undefined factors with potentially undesirable effects. In addition, a number of studies indicated that in contrast to mouse ES cells, LIF was not sufficient to support human ES cell pluripotency in the absence of a MEF layer, and neither was activation of downstream STAT3 (Thomson et al. 1998; Reubinoff et al. 2000; Daheron et al. 2004; Humphrey et al. 2004). Consequently, alternative signalling pathways would have to be modulated to support human ES cell

maintenance in serum- and feeder-free conditions. A comparison of signalling pathways in mouse and human ES cells is shown in Figure B.

FGF signalling emerged as a putative pathway of interest in experiments to clonally culture human ES cells (Amit et al. 2000). One caveat of the original derivation experiments was the passaging of ICM clumps during derivation, rather than clonal expansion from single cells, as despite the homogenous appearance of the resulting cultures there remained the possibility for heterogeneity. However, subsequent experiments demonstrated that H9 ES cells lines could be clonally expanded without loss of pluripotency, though at a lower cloning efficiency than mouse ES cells (Amit et al. 2000). Clonal lines retained a normal karyotype, cell surface marker expression and contributed to teratomas as before. Interestingly, using knockout serum replacement (KOSR) (Price et al. 1998) in place of serum greatly increased cloning efficiency. Furthermore, addition of 4 ng/ml FGF2 to the culture medium resulted in tighter colony morphology, and was required for continuous undifferentiated proliferation in serum-free media (Amit et al. 2000), suggesting a link to pluripotency maintenance. However, even in the presence of FGF human ES cells were still lost to differentiation in the absence of the MEF layer.

Interestingly, human ES cell lines could be maintained on the basement membrane substrates Matrigel and laminin in MEF-conditioned medium (MEF-CM) supplemented with 4 ng/ml of FGF2 (Xu et al. 2001). This suggested that the missing component was secreted by the MEF layer, rather than provided by cell-cell contact. Human ES cells in unconditioned medium alone were subject to high levels of BMP signal-inducing activity, which was diminished in MEF-CM (Xu et al. 2005). As BMP signalling had been linked to human ES cell differentiation (Xu et al. 2002; Pera et al. 2004), this likely partially explained why MEF-CM was required for pluripotency maintenance. However, MEF-conditioned medium could be dispensed with entirely either by adding the BMP antagonist Noggin (500 ng/ml) alongside FGF2 (40 ng/ml), or increasing the concentration of FGF2 to 100 ng/ml in unconditioned medium (Xu et al. 2005).

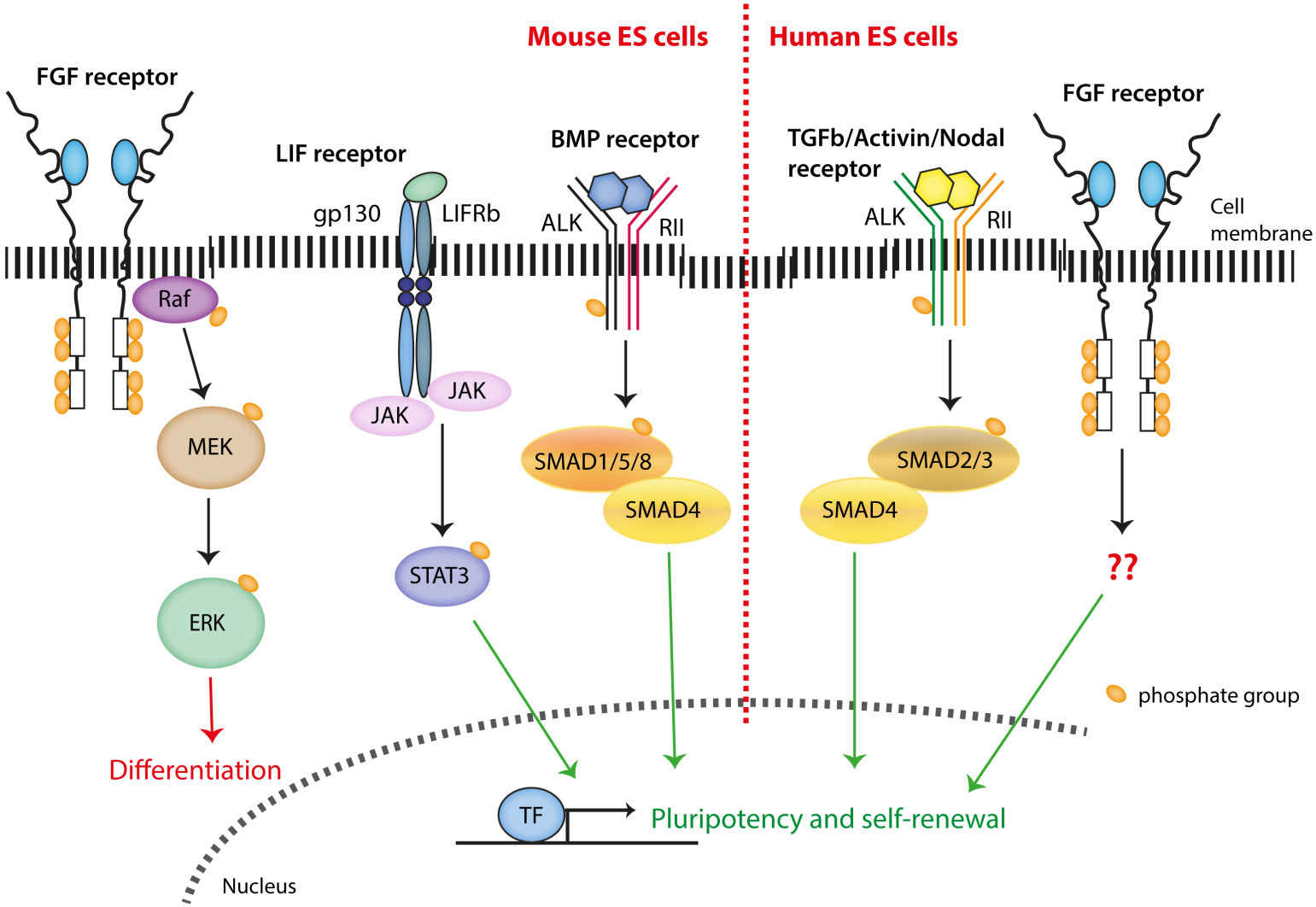


Figure B . Signalling in human and mouse ES cells. Summary of selected signalling pathways required for pluripotency maintenance or that drive differentation in human or mouse ES cells.

Consequently, one function of FGF appeared to be to antagonise the differentiation promoting effects of BMP signalling.

FGF is present in the majority of human ES cell culture media, if not overtly via addition of exogenous ligand, then by culture on MEF layers or in MEF-CM. FGF/Erk inhibition in existing human ES cells affects NANOG expression and promotes neural differentiation (Greber et al. 2010; Greber et al. 2011), consistent with the requirement for FGF to maintain pluripotency. However, as discussed above, Erk inhibition in the human embryo has neither a positive or negative effect on the epiblast compartment (Kuijk et al. 2012; Roode et al. 2012), suggesting that the human epiblast may have distinct downstream signalling requirements.

Core components of the TGFβ/Activin/Nodal signalling pathway have also been shown to be expressed in human ES cells, including the ligand Nodal and the receptors ALK4 and ACVR2B (Vallier et al. 2004). Human ES cells overexpressing the TGFβ ligand Nodal remained undifferentiated in MEF-CM without FGF2 or serum, and maintained SSEA-4 expression. Furthermore, MEF-CM was also shown to contain Activin A protein, and Activin A transcripts were detected in MEF cells (Beattie et al. 2005; Vallier et al. 2005), indicating that MEF-CM is also a source of TGFβ pathway stimulation. Human ES cells cultured in the presence of the TGFβ/Activin inhibitors follistatin or SB-431542 no longer expressed TRA-1-60, OCT4 or NANOG, while media supplemented with 50 ng/ml Activin A was sufficient to maintain pluripotency (Beattie et al. 2005; James et al. 2005; Vallier et al. 2005).

Altogether, Nodal and Activin were found to be interchangeable and both synergised with FGF in otherwise chemically defined medium to maintain human ES pluripotency (Vallier et al. 2005), as did the ligand TGFβ1 (Amit et al. 2004). Activin/Nodal signalling via SMAD2/3 signalling has since been shown to regulate pluripotency gene expression in human ES and iPS cells, binding directly to the *NANOG* promoter, as well as reinforcing expression of Nodal

signalling components (Besser 2004; James et al. 2005; Vallier et al. 2005; Vallier et al. 2009; Brown et al. 2011). Consequently, the FGF and TGFβ/Activin/Nodal signalling pathways form the cornerstone of classical human ES cell culture conditions, including the commonly used defined medium mTeSR™1 (Ludwig et al. 2006) and its further condensed derivative TeSR™-E8™ (Chen et al. 2011). A summary of selected media compositions is provided in Table 1.

However, it has been suggested that FGF only indirectly promotes human ES cell pluripotency, instead stimulating either the supportive MEF layer (Greber et al. 2007), or fibroblast-like cells differentiated from human ES cells themselves (Bendall et al. 2007), to secrete factors that subsequently promote pluripotency. Several members of the insulin-like growth factor (IGF) family have been detected in MEF-CM (Bendall et al. 2007), and IGF2 has been hypothesized to cooperate with FGF to maintain human ES cell pluripotency, functioning via the IGF1 receptor, IGF1R (Bendall et al. 2007; Wang et al. 2007). Addition of an IGF1R-blocking antibody to human ES cell cultures impeded the expansion of cells expressing the pluripotency-associated marker SSEA3 (Bendall et al. 2007), which could not be rescued by supplementation with exogenous FGF2. Furthermore, decreased levels of phosphorylated Akt were observed following IGF1R blocking (Bendall et al. 2007). Briefly, binding of insulin or IGF ligands to insulin or IGF receptors leads to phosphorylation of receptor substrates IRS1 and IRS2, which in turn phosphorylate and activate phosphoinositide 3-kinase (PI3K) (reviewed in (Siddle 2011)).

PI3K phosphorylates the membrane-bound phospholipid PIP2 to generate active PIP3, which then activates a kinase cascade via phosphoinositide-dependent kinase-1 (PDK1) and its substrates Akt and atypical protein kinase Cs (aPKCs). Downstream Akt substrates include Gsk3β, whose links to mouse ES cell maintenance are discussed above, as well as factors involved in apoptosis and protein synthesis (Siddle 2011). However, constitutive Akt expression maintains mouse and non-human primate ES cell pluripotency independently of Wnt/β-Catenin signalling (Watanabe et al. 2006), suggesting that Akt-mediated negative regulation of Gsk3β is

	Conditioned medium (CM)		Unconditioned medium (UM)	CDM	mTeSR1	TeSR-E8
	Amit et al (2000)	Xu et al (2001)	Xu et al (2005)	Vallier et al (2005)	Ludwig et al (2008)	Chen et al (2011) - E8
Basal medium	KO-DMEM, 80%	KO-DMEM, 80%	DMEM-F/12*, 80%	IMDM*, 50% Ham's F-12~, 50%	DMEM/F-12	DMEM/F-12
Serum / replacement	KOSR~, 20%	KOSR~, 20%	KOSR~, 20%			
Additives	L-Glutamine, 1 mM	L-Glutamine, 1 mM	L-Glutamine, 1 mM		L-Glutamine	
	β-mercaptoethanol, 0.1 mM	β-mercaptoethanol, 0.1 mM	β-mercaptoethanol, 0.1 mM		β-mercaptoethanol, 0.1 mM	
	Non-essential amino acids, 1%	Non-essential amino acids, 1%	Non-essential amino acids, 1%			
	~Ascorbic acid 2-phosphate, 64 μg/ml	~Ascorbic acid 2-phosphate, 64 μg/ml	~Ascorbic acid 2-phosphate, 64 μg/ml		L-Ascorbic acid	L-Ascorbic acid 2-phosphate magnesium, 64 μg/ml
	~Sodium selenite, 14 ng/ml	~Sodium selenite, 14 ng/ml	~Sodium selenite, 14 ng/ml	*Sodium selenite, 17 ng/ml	Selenium	Sodium selenium, 14 ng/ml
			*Sodium pyruvate, 55 μg/ml	*~Sodium pyruvate, 110 μg/ml	Sodium pyruvate	Sodium pyruvate, 543 μg/ml
					Lithium chloride	
					GABA	
					Pipecolic acid	
	~Reduced glutathione, 2 μg/ml	~Reduced glutathione, 2 μg/ml	~Reduced glutathione, 2 μg/ml		Glutathione	
	~Thiamine, 6.6 μg/ml	~Thiamine, 6.6 μg/ml	~Thiamine, 6.6 μg/ml		Thiamine	
					Trace elements	
Growth factors	FGF2, 4 ng/ml	FGF2, 8 ng/ml	FGF2, 40 ng/ml	FGF2, 12 ng/ml	FGF2, 100 ng/ml	FGF2, 100 ng/ml
			Noggin, 0.5 μg/ml			
				Activin, 10 ng/ml	TGFβ1, 23.5 pM // Nodal	TGFβ1, 2 ng/ml // Nodal, 100 ng/ml
Additional factors	~Insulin, 20 μg/ml	~Insulin, 20 μg/ml	~Insulin, 20 μg/ml	Insulin, 7 μg/ml	Insulin	Insulin, 19.4 μg/ml
	~Transferrin, 11 μg/ml	~Transferrin, 11 μg/ml	~Transferrin, 11 μg/ml	Transferrin, 15 μg/ml	Transferrin (Human Holo)	Transferrin, 10.7 μg/ml
				Monothioglycerol, 450 μM		
				BSA fraction V, 5 mg/ml	BSA	
				Chemically defined lipid, 1%	Chemically defined lipid	
Support	MEF layer	Matrigel (1:20 dilution) Laminin (human, 20 μg/ml)	Matrigel	FBS-coated plates (2x PBS rinse)	Matrigel Laminin (human, 5 μg/cm2)	Matrigel Vitronectin
Passaging	Collagenase, 1 mg/ml	Collagenase	Dispase, 2 mg/ml		Dispase	EDTA
Notes		MEF-conditioned media				

Expected concentrations for KOSR constituents were extrapolated from the preferred embodiment of the supplement (Price et al., 1998)

Table 1. Comparison of selected defined media formulations for human ES cell culture. Abbreviations: Albumax - bovine serum albumin supplement; BSA - Bovine serum albumin; CDM - chemically defined medium (adapted from Wiles and Johansson, 1995); DMEM-F/12 - Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12; EDTA - Ethylenediaminetetraacetic acid; FBS - Foetal bovine serum; FGF2 - Fibroblast growth factor 2; GABA - gamma-Aminobutyric acid; IMDM - Iscove's Modified Dulbecco's Medium; KO-DMEM - Knockout DMEM (DMEM + KOSR); KOSR - knockout serum replacement; MEF - mouse embryonic fibroblast; TGFβ1- transforming growth factor beta 1

not the sole mechanism of promoting ES cell maintenance. Interestingly, PI3K/Akt signalling can also be activated by FGF (reviewed in (Ornitz and Itoh 2015)), suggesting that FGF also may function via Erk-independent pathways to promote ES cell maintenance. Furthermore, the Grb2 adaptor protein linked to Ras/MAPK signalling can also be activated by insulin receptor substrate Shc (Siddle 2011), suggesting potential crosstalk between these two pathways. Altogether, this indicates that additional pathways may be relevant in human ES cell maintenance.

Searching for the human naïve state

Similar to the mouse, human ES cells require OCT4, NANOG and SOX2 to maintain pluripotency (Wang et al. 2012). Furthermore, induction of c-MYC and the pluripotency factors, OCT4, SOX2 and KLF4 in ES cell culture conditions is also sufficient to reprogram human fibroblasts and other somatic cells to iPS cells (Takahashi and Yamanaka 2006; Yu et al. 2007; Park et al. 2008). However, human and mouse ES cells have distinct molecular signatures (Sato et al. 2004), and human ES cells lack the expression of pluripotency factors associated with the mouse preimplantation Epi such as TBX3 and DAX1, instead expressing the mouse postimplantation marker OTX2 (Loh et al. 2015). Consequently, classical human ES cells are thought to instead represent a later stage of development than their mouse ES cell counterparts, even though both are derived from the preimplantation Epi.

This hypothesis is seemingly supported by the existence of mouse epiblast stem cells (EpiSCs), which are derived from the postimplantation mouse Epi (Brons et al. 2007; Tesar et al. 2007). Human ES cell derivation conditions were tested on pre- and postimplantation mouse embryos, and while the preimplantation ICM underwent rapid differentiation, postimplantation Epi layers formed compact colonies of cells expressing Oct4 and Nanog (Brons et al. 2007). EpiSCs grow in flat epithelial-like monolayers and rely on FGF and Activin/Nodal signalling rather than LIF to maintain pluripotency (Brons et al. 2007; Tesar et al. 2007), similar to human ES cells and

distinct from mouse ES cells. Furthermore, with the exception of Nanog, EpiSCs do not express naïve pluripotency markers, and Oct4 expression in EpiSCs is regulated by the proximal enhancer shown to be required for postimplantation, rather than preimplantation, specific expression in the mouse embryo (Tesar et al. 2007; Han et al. 2010). Although EpiSCs have been successfully derived from a range of developmental stages from E5.5 up to E8.5 (Kojima et al. 2014a), and even from E3.5 preimplantation mouse embryos (Najm et al. 2011), they exhibit a common transcriptomic identity corresponding to the anterior primitive streak (Kojima et al. 2014a). Both human ES cells and mouse EpiSCs do not incorporate into mouse blastocysts to form chimeras (Huang et al. 2012; Masaki et al. 2015), though EpiSCs can contribute to post-implantation embryos (Huang et al. 2012; Kojima et al. 2014a) and human ES cells have recently been shown to contribute to later stage mouse embryos (Mascetti and Pedersen 2016).

Interestingly, if iPS cell reprogramming with the Yamanaka factors is performed in EpiSC culture conditions, the resulting cells exhibit EpiSC morphology and characteristics (Han et al. 2011), indicating the culture environment does have a significant impact on cell identity. When EpiSCs are cultured in the 2i+LIF conditions that would support naïve mouse ES cells, a very small proportion of cells revert to an earlier stem cell identity, but ectopic expression of mouse ES cell genes such as *Klf2, Klf4* or *Nr5a2* is required to increase this in a meaningful way (Guo et al. 2009; Hall et al. 2009; Hanna et al. 2010). Nevertheless, several putative human ES cell culture conditions have recently been investigated, in an attempt to derive stem cells that more closely resemble "ground state" naïve mouse ES cells (Ying et al. 2008).

Studies attempting to derive naïve human ES cells utilise a combination of growth factors and inhibitors to modulate various signalling pathways (Hanna et al. 2010; Chan et al. 2013; Gafni et al. 2013; Takashima et al. 2014; Theunissen et al. 2014; Valamehr et al. 2014; Ware et al. 2014), as well as ectopic transgene expression of *KLF2* and *NANOG* (Takashima et al. 2014; Theunissen et al. 2014). A summary of selected media formulations is shown in Table 2. Putative conditions

	Standard culture conditions	"Naive" human ES cell protocols					
	Vallier (2005)	**Wang (2011)**	**Chan (2013)**	**Gafni (2013)**	**Takashima (2014)**	**Theunissen (2014)**	**Valamehr (2014)**	**Ware (2014)**
Inhibitors		MEKi	MEKi	MEKi	MEKi	MEKi	MEKi	MEKi
		GSK3i	GSK3i	GSK3i	GSK3i	GSK3i	GSK3i	GSK3i
				JNKi				
				P38i				
				PKCi				
				ROCKi		ROCKi	ROCKi	
			BMPi					
						BRAFi		
						SRCi		
					PKCi (Gö6983)			
Growth factors	bFGF		bFGF	bFGF		bFGF**	bFGF	bFGF
	Activin		Activin	TGFβ		Activin		
		hLIF	hLIF	hLIF		hLIF	hLIF	
Base	20% KSR (CDM*)	20% KSR	mTeSR1	Albumax+N2 or 20% KSR	N2B27	N2B27	20% KSR	20% KSR
Support	MEFs (Matrigel*)	STO feeders		MEFs	MEFs	MEFs		MEFs

Table 2: A comparison of putative naive human ES cell culture conditions. A summary of inhibitors, growth factor supplements, basal media and supporting layers used to maintain human ES cells. Abbreviations: Albumax - bovine serum albumin supplement; bFGF - basic FGF (FGF2); BMPi - bone morphogenic protein signalling inhibitor, Dorsomorphin; BRAFi - B-RAF inhibitor SB590885; CDM - chemically defined medium (Wiles and Johansson, 1995); GSK3i- GSK3 inhibitor BIO or CHIR99021; hLIF - human leuakemia inhibitory factor; JNKi - c-Jun N-terminal kinase inhibitor, SP600125; 20% KSR - knockout serum replacement medium containing 20% serum, see Methods for detailed composition; MEFs - mouse embryonic fibroblast feeder cells; MEKi - MEK inhibitor PD0329501; mTeSR1 - serum-free culture medium; N2 - defined serum-free supplement; N2B27 - defined culture medium; p38i - p38 MAPK inhibitor, SB203580; PKCi - protein kinase C (PKC) inhibitor, Gö6983; ROCKi - Rho-associated protein kinase inhibitor Y-27632; SRCi - SRC inhibitor WH-4-023; STO feeders - mouse embryonic fibroblast feeder cells; TGFβ - transforming growth factor beta. * - These conditions apply if grown in MEF-free conditions. ** - bFGF not added directly but likely to be present due to MEF layer. Table adapted from Theunissen et al., 2014, *Cell Stem Cell*, (doi: 10.1016/-j.stem.2014.07.002), with permission under the terms of the Creative Commons Attribution License (CC BY).

often include the Mek and Gsk3β inhibitors used in mouse 2i+LIF media, however culture in 2i+LIF alone is unable to support the self-renewal of human ES cells (Hanna et al. 2010). In contrast, neither LIF nor constitutive STAT3 expression is sufficient to maintain pluripotency of existing human ES cells (Thomson et al. 1998; Humphrey et al. 2004). This may impact the conclusions drawn from naïve protocols in human ES cells, as the benchmark against which these cells are assessed relies heavily on conclusions drawn from mouse ground state pluripotency. Transcription factors such as *KLF2* and *ESRRB* are not appreciably expressed in the human Epi (Blakeley et al. 2015), though both are thought to be essential for maintaining ground state pluripotency in the mouse (Martello et al. 2013; Yeo et al. 2014). This complicates the use of mouse naïve factors either as a driving force towards or as an indicator of a more naïve human pluripotent state. Despite this, gene expression patterns in naïve human ES cells are more similar to the human Epi than to naïve mouse ES cells (Huang et al. 2014), suggesting that perhaps these conditions may inadvertently intersect with human pluripotency maintenance networks. Some studies also demonstrate a switch from *OCT4* expression from the distal enhancer in naïve human ES cells, rather than the proximal enhancer active in primed cells (Gafni et al. 2013; Takashima et al. 2014; Theunissen et al. 2014; Ware et al. 2014), which is also based on naïve pluripotency dynamics in the mouse (Yeom et al. 1996; Choi et al. 2016). However, this assumes similar modes of upstream regulation of the OCT4 enhancer and promoter region in the human embryo, and although the sequences are conserved in the mouse and human genes (Nordhoff et al. 2001), there is currently no further data to support this hypothesis.

Further analysis directly in human embryos may reveal human-specific signalling networks that could be modulated, rather than relying on those characterised in rodent models. In addition, recent gene expression studies in the human embryo may also provide new benchmarks for assessing the relationship of naïve conditions to *in vivo* pluripotency, such as human Epi-specific transcription factors like KLF17 (Blakeley et al. 2015) or additional molecular criteria such as

transposon signatures (Theunissen et al. 2016). Indeed, upregulated *KLF17* transcript or KLF17 protein expression can be detected in certain naïve human ES cell cultures (Blakeley et al. 2015; Guo et al. 2016; Collier et al. 2017). Increased genome-wide DNA demethylation following adaptation to naïve culture conditions (Gafni et al. 2013; Guo et al. 2016; Theunissen et al. 2016) also correlates with the hypomethylated status observed in the human ICM (Guo et al. 2014; Smith et al. 2014), which also happens to characterize naïve pluripotency in mouse ES cells (Leitch et al. 2013).

Nevertheless, human ES cells remain somewhat distinct from the human Epi (Yan et al. 2013; Blakeley et al. 2015), suggesting that current *in vitro* culture conditions, naïve or otherwise, do not fully reflect the conditions required for pluripotency maintenance *in vivo*. Although naïve human ES cells are more similar to the human embryo than their primed counterparts (Yan et al. 2013; Blakeley et al. 2015), they do not necessarily completely resemble a specific developmental stage. This is in contrast to the mouse system where mouse naïve or primed ES cells, and mouse EpiSCs show similar gene expression profiles to, and cluster with, embryonic cell types (Boroviak et al. 2014; Boroviak et al. 2015). Additionally, certain cell surface markers have recently been shown to accurately delineate naïve human ES cell populations, but do not necessarily exhibit an Epi-restricted expression pattern in the human embryo (Collier et al. 2017). It is likely that further refinement is required to derive human ES cells that more closely resemble the human embryo. Nevertheless, given that naïve human cells adopt embryo-associated characteristics, such as for example, global demethylation, studying the transition between primed and naïve cells presents a useful model to better understand how these processes occur. Furthermore, the various naïve cells share a core transcriptional network despite their different culture systems (Huang et al. 2014). It would be interesting to further analyse these genes to determine if they lend insight into the fundamental underpinnings of the human pluripotent state. Characteristics of mouse ES cells and EpiSCs compared to conventional and naïve human ES cells are shown in Table 3.

	Mouse			Human	
	ES cell	**Naïve ES cell**	**EpiSC**	**ES cell**	**Naïve ES cell**
Embryonic origin	Pre-implantation ICM (E3.5)	Pre-implantation ICM (E3.5)	Post-implantation Epi (E5.5 - E8.0)	Pre-implantation ICM (E6)	Pre-implantation ICM (E6)
In vitro culture					
Media composition	Serum (BMP) + LIF	2i + LIF	CDM + FGF + Activin	CDM + FGF + Activin [a]	Various [a]
Colony morphology	Dome-shaped	Dome-shaped	Flattened, epithelial	Flattened, epithelial	Dome-shaped
Pluripotency criteria					
In vitro differentiation	Yes	Yes	Yes	Yes	Yes
Teratoma formation	Yes	Yes	Yes	Yes	Yes
Blastocyst integration	Yes	Yes	No	- [b]	- [b]
Post-implantation integration	No	No	Yes	Yes	-
Germline transmission	Yes	Yes	No	- [b]	- [b]
Gene expression					
Core pluripotency factors (Nanog, Oct4, Sox2)	Yes	Yes	Yes	Yes	Yes
Mouse naïve factors (Esrrb, Klf2)	Yes	Yes	Low	No	Yes
Human Epi-specific factors (KLF17, ARGFX)	No	No	No	No	Yes [c]
Dominant Oct4/OCT4 enhancer	Distal	Distal	Proximal	Proximal	Distal
Epigenetic profile					
Global DNA hypomethylation	No	Yes	No	No	Yes
H3K27me3 on developmental regulators	No	Yes	No	No	Yes
X chromosome inactivation	No	No	Yes	Yes	No

Table adapted from De Los Angeles, A., et al. (2015), Nature 525(7570): 469-478, and Weinberger, L., et al. (2016), Nat Rev Mol Cell Biol 17(3): 155-169.

[a] See Table 2 for more detailed conditions for standard and naïve human ES cell culture. Naïve characteristics listed here largely reflect culture parameters from Takashima (2014) and Theunissen (2014)

[b] Not applicable due to ethical considerations; cells also generally integrate poorly into inter-species chimeras

[c] Upregulated in some but not all existing naïve culture conditions - see text for details

Table 3: A comparison of selected defining criteria of various mouse and human stem cell types.

Summary

Signalling and gene regulatory networks cooperate to determine lineage segregation in the embryo, and modulating signalling *in vitro* allows for stem cell populations to be established from these lineages. FGF signalling in particular destabilises pluripotency in the Epi or in mouse ES cells, but promotes the PrE lineage in the mouse preimplantation embryo. Precisely how FGF functions to drive these alternative outcomes is only just beginning to be elucidated. Curiously, FGF is instead routinely used to maintain pluripotency in human ES cells, and in EpiSCs derived from post-implantation mouse Epi. Human ES cells generally more closely resemble EpiSCs than mouse ES cells in their morphology and signalling requirements, despite being derived from the preimplantation embryo. This would seemingly suggest either a species-specific discrepancy related to FGF signalling in the preimplantation embryo, or differences in the downstream effectors of FGF in these alternative contexts.

Consequently, the work described in this study aimed to address two main areas. Firstly, to interrogate the downstream mechanics of FGF/Erk signalling in mouse ES cells, where the effect of FGF signalling correlates with that in the embryo. Determining how signalling functions to destabilise the existing pluripotency network could provide insights into how FGF preferentially promotes a PrE identity during lineage specification in the ICM. An Erk2-Venus fluorescent reporter mouse ES cell line was used to investigate Erk localisation dynamics in response to FGF stimulation, and how this is related to lineage segregation *in vivo* and ES cell differentiation *in vitro*. In addition, a number of transgenic inducible mouse ES cells were characterised in order to investigate how lineage-specifying transcription factors integrate with FGF signalling to disrupt pluripotency regulation.

The second aim was to investigate FGF signalling more thoroughly in the human ICM, making use of access to donated human embryos that were surplus to family building, in an attempt to reconcile the divergent outcomes of FGF/Erk inhibition in human ES cells compared to the

human embryo. Finally, a recently acquired single cell transcriptomics dataset was explored to characterise human Epi-specific gene expression and more comprehensively determine conditions that would best support pluripotency maintenance.

Materials and Methods

1. Mouse cell culture

1.1. Routine maintenance of mouse ES cells, XEN cells and neural stem cells

Mouse ES cells were generally maintained on mouse embryonic fibroblast (MEF)-coated pre-gelatinized tissue culture plates (Corning) in serum and 10 ng/ml of LIF (KSR+LIF). Recombinant LIF was produced by Marko Hyvonen (University of Cambridge, Department of Biochemistry). MEF plates were prepared 24 hours in advance using MEF medium. Full KSR and MEF media compositions are listed in the Reagents section below. Mouse ES ells were passaged every 2-3 days with 0.05% Trypsin-EDTA (Sigma).

For experiments carried out in the absence of exogenous FGF, mouse ES cells were cultured in serum-free neurobasal medium (NDiff 227, Cellartis/Takara Bio) supplemented with 1 μM of Erk inhibitor PD0325901 (Tocris), 3 μM of Gsk3 inhibitor CHIR99021 (Axon), and 10 ng/ml of LIF (2i+LIF conditions, (Ying et al. 2008). Cells were cultured on gelatine-coated tissue culture plates in 2i+LIF for at least 3 passages prior to any experimental setup.

ZHBTc4 Oct4 conditional knockout mouse ES cells (Niwa et al. 2000) were routinely cultured as above, and treated with 1 mg/mL of doxycycline as required to generate Oct4-null cells. ZHBTc4 mouse ES cell lines were kindly supplied by Austin Smith.

Mouse embryo-derived XEN (eXEN) cells, and transcription factor-induced XEN (iXEN) cells generated in this study, were routinely maintained on gelatine-coated tissue culture plates in XEN maintenance media (adapted from (Kunath et al. 2005). The full XEN media composition is listed in the Reagents section below. Mouse eXEN cells were kindly supplied by Janet Rossant.

The NS5 neural stem cell line (Conti et al. 2005) was cultured in Euromed-N media (EuroClone) supplemented with L-glutamine, laminin, N2, BSA, FGF2 and EGF (R&D or Life Technologies). Cells were routinely passaged using Accutase (Life Technologies). NS5 cells were kindly supplied by Zachary Gaber and Francois Guillemot.

Mouse cells were frozen down in a minimal volume of 10% DMSO in serum as standard. When thawed, cells were resuspended in the relevant culture media and washed twice to remove traces of freezing medium (cells were pelleted by centrifugation, supernatant was aspirated and replaced with fresh media) before plating for culture.

1.2. Generation of transgenic mouse ES cell lines

Erk2-Venus mouse ES cell lines were generated in collaboration with Russell Ernst. R1 mouse ES cells (Nagy et al. 1993) were electroporated with a pCAGGs vector containing a C-terminal targeting construct incorporating part of Exon 8 of the *Erk2* allele followed by an in-frame *Venus* tag, as well as a hygromycin resistance sequence separated from the *Venus* sequence by an T2A peptide. Successfully targeted ES cell colonies were selected using hygromycin (50-100 μg/mL).

Doxycycline-inducible cell lines were generated by Lily Cho and Kathy Niakan using KH2 mouse ES cells (Beard et al. 2006), previously modified to include an M2-rtTA tetracycline-responsive transactivator under the control of the *Rosa26* promoter, plus a tetracycline operator minimal promoter incorporated into the *Col1a* locus (Hochedlinger et al. 2005; Beard et al. 2006). *Gata4, Gata6, Hnf4a, Foxa3, Sox7* and *Sox17* cDNA sequences were PCR-amplified from XEN cell lines and integrated into a pgkATGfrt vector that also conferred hygromycin resistance, then co-electroporated into KH2 cells with a pCAGGS-FLPe transient vector to facilitate recombination. Successfully targeted colonies were selected using puromycin (1 μg/ml) and hygromycin (50-100 μg/mL).

Fgf4$^{-/-}$ mouse ES cells (Kang et al. 2013) were kindly supplied by Anna-Katerina Hadjantonakis. Lily Cho replicated the *Gata6*-inducible system in these cells by integrating the M2-rtTA tetracycline-responsive transactivator into the endogenous *Rosa26* locus followed by targeting with the frt-neomycin/hygromycin-pA vector into the *Col1A* locus. Targeted cells were selected with puromycin (1 μg/ml) and neomycin (200 μg/mL). Double targeted *Fgf4*$^{-/-}$ cells were then electroporated with the pgkATG-Gata6 and pCAGGS-FLPe vectors, and successfully targeted colonies were selected using hygromycin (50-100 μg/mL).

1.3. Doxycycline induction of transgene expression

Transgene expression was induced 24 hours after mouse ES cells were plated, with doxycycline added at a final concentration of 1 mg/mL to serum and LIF (KSR+LIF) media, LIF-deficient media or media supplemented with FGF and Heparin, as required. Recombinant FGF4 (R&D) was used at 1 μg/ml, with Heparin at 1 μg/ml. Transgene expression was induced for up to 6 days (144 hours) and samples collected at defined timepoints as required.

Gata6-inducible *Fgf4*$^{-/-}$ mouse ES cells were cultured in 2i+LIF for at least 3 passages prior to experimental setup, and doxycycline induction was carried out in serum-free neurobasal medium on gelatin. For *Pou5f1* knockdown experiments, shRNAs were introduced into *Gata6*-inducible mouse ES cells concomitant with doxycycline induction in KSR+LIF. Five distinct constructs were used over three experiments.

For *Gata6* pulse chase experiments, transgene expression was induced with doxycycline-containing KSR+LIF for the required pulse period. Media was then aspirated and cells washed twice with pluripotency media, and then cultured in fresh pluripotency media for the 48-hour chase.

1.4. *GATA6* lentiviral induction

Lentiviral packaging in HEK293T cells was carried out by Ignacio del Valle, using either Lipofectamine 3000 (Life Technologies) or X-fect transfection reagent (Clontech). HEK293T cells were kindly provided by Lucas Baltussen and Sila Ultanir. Packaging plasmids were co-transfected with a plasmid containing human HA-tagged *GATA6* and a puromycin resistance gene under control of an *EF1a* promoter (AMSbio). Lentiviral supernatants were obtained and concentrated by ultracentrifugation.

Mouse ES or NS5 cells were plated on gelatin-coated tissue culture plates 24 hours before transduction so as to be 70-80% confluent at the time of transduction. Cells were transduced with the *HA-GATA6* lentivirus for 12 hours, and then switched to media containing 1 μg/mL puromycin (Sigma) for selection. For ZHBTc4 Oct4 conditional knockout mouse ES cells, after 12 hours cells were switched to media supplemented with 1 mg/ml doxycycline as well as 1 μg/ml puromycin.

1.5. *Esrrb* overexpression in *Gata6*-induced cells

Gata6-inducible mouse ES cells were plated on gelatin-coated tissue culture plates 24 hours before transfection so as to be 70-80% confluent at the time of transfection. pCAGIP-Myc-Esrrb (Uranishi et al. 2013) or PpyCAG-FLAG-Esrrb (van den Berg et al. 2008) plasmids were introduced using the X-fect transfection reagent (Clontech), following the manufacturer's instructions. Cells were supplemented with 1 mg/mL doxycycline concomitant with transfection, and samples collected and analysed 24 hours later.

2. Mouse embryo harvest and culture

6 weeks-old female Parkes or B5/EGFP mice were naturally mated. After detection of a vaginal plug (E0.5), females were sacrificed and oviducts dissected and removed to a petri dish containing drops of M2 media (made in-house, according to Jackson Laboratories protocol) at

room temperature. Piercing the ampulla released the clutch of zygotes into a drop of hyaluronidase solution, where they remained for a few minutes in order to remove the cumulus cells. Zygotes were then collected and washed through several drops of M2 media before being transferred to a micro drop culture plate containing equilibrated KSOM media (made in-house, according to Jackson Laboratories protocol), overlaid with mineral oil (Sigma). Zygotes were washed through several drops of KSOM media, then transferred to fresh drops of equilibrated KSOM media and cultured at 37°C, 5% CO_2, to specific stages of development.

In instances where mouse embryos were used to test growth factor toxicity prior to human embryo culture, embryos were cultured in individual drops of Global media (Life Global) supplemented with growth factors or inhibitors as required. For further analysis, embryos were routinely fixed in 4% paraformaldehyde (Sigma) for 1 hour on ice.

2.1. Chimera generation following *Gata6* pulse-chase induction

For chimera generation, cells were injected into the blastocoel cavity of E2.5 – E3.5 B5/EGFP blastocysts (Hadjantonakis et al. 1998). To investigate contribution to the primitive endoderm, single cells were injected and embryos cultured in KSOM till E4.5, then fixed for further analysis. To investigate contribution to ExEn-derived tissues, 3 to 5 cells were injected per embryo and embryos transferred to the uterus of E2.5 vas plugged females. Kathy Niakan then performed dissections for embryo retrieval between E5.5 and E6.5. Cell injections and blastocyst transfer to recipient females were carried out by the Genetic Manipulation Services (GEMS) core facility.

3. Human embryo and ES cell culture conditions

3.1. Human embryo thaw and culture

Cryopreserved human embryos surplus to family building were donated to the research project by informed consent under the UK Human Fertilisation and Authority Licence number R0162.

Embryos were thawed according to the requirements of the embryo developmental stage, and of the particular method and devices used for initial cryopreservation. Vitrified embryos were thawed using Vit Kit®-Thaw (Irvine Scientific) reagents. Slow frozen embryos were thawed using Quinn's Advantage Thaw Kit (Sage) or BlastThaw™ (Origio) reagents. Embryos frozen in straws were thawed by quickly transferring the contents of the straw from liquid nitrogen onto a petri dish for embryo retrieval, or directly into thaw solution. Embryos frozen in Cryopette® devices were first thawed for 3 seconds in a 37°C water bath before transfer. Embryos frozen in glass ampoules were thawed completely in a 37°C water bath before transfer, after the top of the vial was removed under liquid nitrogen.

Embryos were routinely cultured in micro drops of Global Media (Life Global) supplemented with 5 mg/mL Life Global Protein Supplement and overlaid with mineral oil (LiteOil, Life Global). Drops were pre-equilibrated overnight in an incubator at 37°C, 5% CO_2. For growth factor treatment, these conditions were supplemented with the following recombinant human proteins: FGF2 (R&D) at 100 ng/ml, or at 1 μg/ml with 1 μg/ml heparin; IGF1 (R&D) at 1.7 nM or 17 nM; Activin A (R&D) at 10 ng or 50 ng, as indicated in the text. For TGFβ/Nodal signalling inhibition, media was supplemented with 40 μM SB-431542 (Sigma).

3.2. Routine maintenance of human ES cells

H1 and H9 human ES cells (WiCell) were generally cultured on Matrigel-coated (BD Biosciences) tissue culture plates in mTeSR1 media (Stemcell Technologies). Human ES cells were passaged as clumps using ReLeSR (Stemcell Technologies). Briefly, ReLeSR was added to wells for 30 seconds and removed, then plates were incubated for 5 minutes at 37°C, quenched with mTeSR and lightly tapped to dislodge clumps of the desired size. H1 and H9 cells were also adapted to plates coated with 1 μg/cm^2 laminin-511 (Biolamina, Takara) in mTeSR1. Cells on laminin-511 were also passaged with ReLeSR, but with a 7 minute incubation at 37°C.

Human cells were frozen down in a minimal volume of 10% DMSO in serum as standard. When thawed, cells were resuspended in the relevant culture media and washed thrice to remove traces of freezing medium (cells were pelleted by centrifugation, supernatant was aspirated and replaced with fresh media) before plating for culture.

3.3. Doxycycline inducible *GATA6* lentiviral induction of human ES cells

Doxycycline-inducible *GATA6*-expressing human ES cells were generated by Ignacio del Valle using the Lenti-X Tet-On 3G inducible expression system (Clontech), according to the manufacturers protocol. Briefly, the HA-tagged human GATA6 coding sequence was sub-cloned into the pLVX-TRE3G vector provided. HEK 293T cells were then transfected with the resulting pLVX-TRE3G-hGATA6-HA, or with a pLVX-Tet3G vector, using the X-fect reagent (Clontech). Lentiviral supernatants were concentrated using the Lenti-X Concentrator (Clontech). H9 human ES cells were then sequentially transduced with the pLVX-Tet3G and the pLVX-TRE3G-hGATA6-HA lentiviruses, with selection with G418 (250 μg/ml) for at least four passages prior to the second transduction. Human *GATA6* expression was induced with doxycycline at a concentration of 1 mg/ml.

4. Molecular analysis

4.1. Western Blotting

Whole cell protein was extracted with CelLytic M reagent (Sigma) supplemented with proteinase and phosphatase inhibitors (Roche). 30 μg of protein per sample was resolved on 12% SDS-PAGE gels and transferred to a PVDF membrane using a BioRad Trans-Blot transfer system (BioRad). Membranes were blocked in TBS and 0.1% Tween (Sigma) with 5% skimmed milk, or 5% BSA (Sigma) for detection of phosphorylated proteins, and incubated with primary antibody at 4°C overnight. Following washes in TBS 0.1% Tween, membranes were incubated with secondary antibody in 5% milk or 5% BSA for one hour at room temperature. Proteins were visualized using the Pierce ECL Western Blotting Substrate (Thermo) with film exposure.

4.2. Immunofluorescence and imaging

Embryos and ES cells were fixed in 4% paraformaldehyde at 4°C for 1 hour or overnight respectively, then permeabilized with PBS 0.5% Tween for 20 minutes and blocked with 10% FBS (Life Technologies) in PBS 0.1% Tween for 1 hour. Primary antibodies were diluted at 1:500 in blocking solution and samples incubated at 4°C rotating overnight. Secondary antibodies were diluted at 1:300 in blocking solution and samples incubated for 1 hour at room temperature, then washed and covered with PBS 0.1% Tween containing DAPI Vectashield mounting medium (Vector Labs).

Embryos were imaged on a Leica SP5 inverted confocal microscope (Leica Microsystems) at a z-section thickness of 3μm or 2 μm for human or mouse embryos respectively. ES cells were imaged using an Olympus 1X71 microscope with Cell^F software (Olympus Corporation, Tokyo, Japan), Zeiss Axiovert 200M microscope with AxioVision Rel 4.7 software (Carl Zeiss Ltd., Jena, Germany).

MINS 1.3 software was used to analyse immunofluorescence in embryos (http://katlab-tools.org/) (Lou et al. 2014). Confocal stacks in .tif format were loaded into the MINS pipeline for automated nuclear detection and segmentation to determine the number of cells in each embryo. The MINS segmentation output was manually checked for appropriate segmentation and mitotic nuclei were removed from the analysis. Data were subsequently plotted using GraphPad Prism version 7 (GraphPad Software, La Jolla, CA).

4.3. Quantitative RT-PCR (qRT-PCR)

RNA was isolated using TRI Reagent (Sigma) and DNaseI treated (Ambion). cDNA was synthesized using a Maxima First Strand *cDNA* Synthesis Kit (Fermentas). qRT-PCR was performed using Quantace Sensimix on an Applied Biosystems 7500 machine (Life Technologies

Corporation, California, USA). Primer pairs were previously published (Brown et al., 2010a; Fujikura et al., 2002; Molkentin et al., 1997; Niwa et al., 2005) or designed using Primer3 software. Primers are listed in the Reagents section.

4.4. Flow cytometry

Cells were dissociated using 0.05% Trypsin-EDTA, then washed twice and resuspended in FACS fixing buffer (1% paraformaldehyde, PBS 0.01% Tween) for 20 minutes on ice. Cells were washed twice with permeabilisation buffer (5% FBS, PBS 0.05% Tween) then incubated with Nanog primary antibody (REC-RCAB0001P, 2B Scientific) and FITC anti-rabbit secondary antibody (A21206, Alexa Fluor) at a 1:250 dilution in permeabilisation buffer for 20 minutes on ice. Cells were washed twice then resuspended in permeabilisation buffer for analysis. Cells were analysed on a BD Fortessa X20 flow cytometer (BD Biosciences). FlowJo software (BD Biosciences) was used to generate histogram overlays and dot plots.

4.5. Microarray analysis

Total RNA was isolated as above and DNaseI treated (Ambion). RNA quality was assessed on a Eukaryote Total RNA Nano Series II (Agilent Technologies, Santa Clara, CA, USA) then processed on an Agilent 2100 Bioanalyzer using the RNA electrophoresis program. All RNA samples were amplified using the Total Prep 96 RNA amplification kit (Ambion). Samples were hybridized to Illumina MouseWG-6_V2 expression BeadChip arrays (Illumina, Inc. California, USA) Biological triplicates were collected for each sample. Details of computational analysis can be found in Wamaitha et al., 2015, Supplemental Experimental Procedures. Data is available via the Gene Expression Omnibus (GSE69321).

4.6. Chromatin immunoprecipitation sequencing (ChIP-seq)

Gata6-inducible mouse ES cells were seeded at 1 x 10^4 cells/cm^2 and treated with 1mg/mL doxycycline for 36 hours prior to harvesting. Immunoprecipitation was performed on 1-2 x10^7

cells as described (Vokes et al. 2007) for three biological replicates versus input samples. Sonication was performed using a Misonix 4000 (28 cycles: 15 seconds on and 45 seconds off at an intensity of 70%) with a Microtip Probe (Misonix, Farmingdale, NY). Antibodies used are listed in the Reagents section. Libraries were prepared using the TruSeq ChIP Sample Preparation Kit and the resulting samples were sequenced using Illumina Genome Analyzer II (Illumina, San Diego, CA). Details of computational analysis can be found in (Wamaitha et al. 2015), Supplemental Experimental Procedures. Data is available via the Gene Expression Omnibus (GSE69322).

4.7. RNA sequencing

RNA was isolated from untransduced and *Gata6*-transduced human ES cells using TRI Reagent (Sigma) and DNaseI treated (Ambion). Libraries were prepared using the TruSeq RNA Library Preparation Kit v2 (Illumina). Quality of total RNA and subsequent cDNA library preparations was assessed on a 2100 Bioanalyzer (Agilent). Libraries were submitted for 50-bp paired-end sequencing on an Illumina HiSeq 2000 (Illumina). Details of computational analysis can be found in (Wamaitha et al. 2015), Supplemental Experimental Procedures. Data is available via the Gene Expression Omnibus (GSE69322).

5. Reagents

5.1 KSR media composition (500 ml)

Component	Supplier	Catalogue Number	Stock Concentration	Final concentration	Volume
Advanced DMEM/F12	Invitrogen	12634-010	-	78%	390 ml
Knockout serum replacer (KOSR)	Invitrogen	10828028	-	20%	100 ml
L-glutamine	Invitrogen	25030-024	200 mM	2 mM (1%)	5 mL
β mercaptoethanol	Sigma	M6250	14.3 M	0.1 mM	3.5 µl
Penicillin/ Streptomycin	Sigma	15140-122	100%	1%	5 ml

5.2 MEF media composition (500 ml)

Component	Supplier	Catalogue Number	Stock Concentration	Final concentration	Volume
Advanced DMEM/F12	Invitrogen	12634-010	-	88%	440 ml
Foetal Bovine Serum (FBS)	Bioserum	S1818	-	10%	50 ml
L-glutamine	Invitrogen	25030-024	200 mM	2 mM (1%)	5 mL
β mercaptoethanol	Sigma	M6250	14.3 M	0.1 mM	3.5 µl
Penicillin/ Streptomycin	Sigma	15140-122	100%	1%	5 ml

5.3. XEN media composition (500 ml)

Component	Supplier	Catalogue Number	Stock Concentration	Final concentration	Volume
RPMI 1640	Invitrogen	61870-010	-	83%	415 ml
Foetal Bovine Serum (FBS)	Bioserum	S1818	-	15%	75 ml
L-glutamine	Invitrogen	25030-024	200 mM	2 mM (1%)	5 mL
β mercaptoethanol	Sigma	M6250	14.3 M	0.1 mM	3.5 µl
Penicillin/ Streptomycin	Sigma	15140-122	100%	1%	5 ml

5.4. Antibodies

Primary Antibody	Supplier	Catalogue Number	Application	Species	Dilution and Reactivity*
Actin	Millipore	MAB1501	WB	Mouse	1:10000
Esrrb	R&D	PP-H6705-00	IHC	Mouse	1:500
FLAG	Sigma	F1804	IHC/WB/ChIP	Mouse	1:500/1:1000/ 9mg per mL
Gata4	Santa Cruz Biotech	SC-9053	IHC/WB	Rabbit	1:500/1:1000
Gata6	R&D	AF1700	IHC/WB	Goat	1:500/1:1000 Mouse
Gata6	Santa Cruz Biotech	SC-9055	IHC	Rabbit	1:500 Human
HA	Roche	11867423001	IHC	Rat	1:500
Total p44/42 MAPK (Erk1/2)	Cell Signalling	4695	WB	Rabbit	1:2000
Phospho p44/42 MAPK	Cell Signalling	9101	WB	Rabbit	1:1000
Nanog	2B Scientific	REC-RCAB0001P	IHC/WB	Rabbit	1:1000 Mouse
Nanog	R&D	AF1977	IHC	Goat	1:500 Human
Oct4	Santa Cruz Biotech	SC-5279	IHC/WB	Mouse	1:500/1:1000
Sox2	R&D	AF2018	WB	Goat	1:1000
Sox7	R&D	MAB2766	IHC	Goat	1:500
Sox17	R&D	AF1924	IHC/WB	Goat	1:500/1:1000

*If not specifically indicated, antibody is reactive in both mouse and human.

Secondary Antibody	Supplier	Catalogue No.	Application	Species	Dilution
Alexa Fluor anti-mouse IgG	Invitrogen	A21202 (488) A21203 (594) A31571 (647)	IHC	Donkey	1:300
Alexa Fluor anti-rabbit IgG	Invitrogen	A21206 (488) A21207 (594) A31573 (647)	IHC	Donkey	1:300
Alexa Fluor anti-goat IgG	Invitrogen	A11055 (488) A11058 (594) A21447 (647)	IHC	Donkey	1:300
Goat IgG (H+L), HRP conjugated	Santa Cruz	SC-2020	WB	Goat	1:20000
Mouse IgG (H+L), HRP conjugated	Cell Signalling	7076	WB	Mouse	1:20000
Rabbit IgG (H+L), HRP conjugated	Cell Signalling	7074	WB	Rabbit	1:20000

5.5. Mouse qRT-PCR primer pairs

Name	Forward Primer	Reverse Primer
Col4a1	GGTCCTGTCTGGAAGAGTTT	AAATACAATGGGAGGGAGAA
Dab2	GGCAACAGGCTGAACCATTAGT	TTGGTGTCGATTTCAGAGTTTAGAT
Esrrb 3'UTR	CCTCCACTTTCCCCTTTCTT	AGGTTTCTAGACTGGCCATG
FLAG	CAAAGACCATGACGGTGATTA	TCGTCATCCTTGTAGTCGATG
Foxa3 3'UTR	GGCTGTGTGCCTCAGGTCGG	CTCCGGTGCTGCCACCAGTG
Gapdh	AATGGAATACGGCTACAGC	GTGCAGCGAACTTTATTG
Gata4 3'UTR	TTCCTGCTCGGACTTGGGAC	TGGACAGGCAGGTGGAGAATAAG
Gata6 3'UTR	ACAGCCCACTTCTGTGTTCCC	GTGGGTTGGTCACGTGGTACAG
V5	CGGTTCGAAGGTAAGCCTAT	ATGGTGATGGTGATGATGA
Hnf4a 3'UTR	CAACGATCACCAAGCAAGAA	CTGGGGTCTTCTCAAGCAAA
Lama1	AGGTCTGCGTTGAGTGTTCTG	CAGTACTATGCCGTCAGCGAT
Nanog	AGCAGATGCAAGAACTCTCCTC	AAGTTGGGTTGGTCCAAGTCT
Pou5f1	AGAAGTCCCAGGACATCAAA	TGCTTTGCATATCTCCTGAA
Prdm14 3'UTR	ATTTGGTTTTGGTGGTGGTG	CCTGGCATTTTCATTGCTCA
Sox2	TAGAGCTAGACTCCGGGCGATGA	TTGCCTTAAACAAGACCACGAAA
Sox7 3'UTR	AACACTCCCGGCCACCCTGA	AAGGAGCGCTGGATCCTGGCT
Sox17 3'UTR	AAGAAACCCTAAACACAAACAGCG	TTTGTGGGAAGTGGGATCAAGAC
Sparc1	AGGGCCTGGATCTTCTTTCTC	CAAATTCTCCCATTTCCACCT

5.6. Human qRT-PCR primer pairs

Name	Forward Primer	Reverse Primer
AFP	ACCATGAAGTGGGTGGAATC	TGGTAGCCAGGTCAGCTAAA
GATA6	TCCACTCGTGTCTGCTTTTG	TCCTAGTCCTGGCTTCTGGA
HA-GATA6	AACACAACCTACAGCCTCAG	GACGTCATAAGGGTAGGCC
NANOG	GCAACCTGAAGACGTGTGAA	CTCGCTGATTAGGCTCCAAC
POU5F1	TATGGGAGCCCTCACTTCAC	CAAAAACCCTGGCACAAACT
SOX2	TTGTTCGATCCCAACTTTCC	ACATGGATTCTCGGCAGACT
SOX7	ACGCCGAGCTCAGCAAGAT	TCCACGTACGGCCTCTTCTG
SOX17	CGCACGGAATTTGAACAGTA	GGATCAGGGACCTGTCACAC

Results and Discussion

1. FGF signalling dynamics and lineage specification in mouse ES cells

The precise mechanism by which ICM cells in the early embryo diverge to favour either a pluripotency or an ExEn gene regulatory network remains unclear. However, activation of FGF signalling has been shown to be required for PrE specification, and derivation of mouse XEN cells requires FGF, though endogenous levels are sufficient (Feldman et al. 1995; Arman et al. 1998; Kunath et al. 2005; Chazaud et al. 2006; Yamanaka et al. 2010; Cho et al. 2012; Grabarek et al. 2012; Kang et al. 2013; Niakan and Eggan 2013). FGF signalling has also been linked to mouse ES cell differentiation (Kunath et al. 2007; Stavridis et al. 2007; Ying et al. 2008; Cho et al. 2012).

A reciprocal FGF receptor-ligand relationship is thought to exist between the PrE and Epi lineages within the mouse ICM (Guo et al. 2010), and this correlates with the effects of FGF signalling modulation in the embryo. FGF stimulation converts all ICM cells to PrE (Chazaud et al. 2006; Yamanaka et al. 2010), likely by destabilising the balance of receptor-enriched PrE versus receptor-deficient Epi. Conversely, mutations in FGF signalling pathway components limit all cells within the ICM to an Epi fate (Feldman et al. 1995; Arman et al. 1998; Cheng et al. 1998). Given that inhibiting Mek phenocopies mutations of upstream FGF signalling components (Nichols et al. 2009; Yamanaka et al. 2010), its downstream target, Erk, is likely to be involved in these lineage decisions.

Previous studies in mouse ES cells indicate that FGF signalling functions via the Raf-Mek-Erk MAPK pathway to disrupt pluripotency (Burdon et al. 1999; Ying et al. 2008). Two Erk homologues, Erk1 and Erk2 are present in mammalian cells, but only Erk2 has an embryonic lethal phenotype (Pages et al. 1999; Hatano et al. 2003), and is thus likely to be the more developmentally relevant kinase. However, it is unclear precisely how Erk phosphorylation

induces these cell fate choices, and whether this is linked to the nuclear localisation dynamics observed in other contexts (Robinson et al. 1998; Brunet et al. 1999; Costa et al. 2006).

In this chapter, I investigate Erk localisation dynamics in mouse ES cells and the E3.5 preimplantation embryo, as well as exploring the relationship between FGF/Erk signalling and the PrE-associated factor Gata6 during ES to XEN cell reprogramming.

1.1. FGF signalling and Erk localisation in mouse ES cells

In alternative contexts, Mek phosphorylates Erk and triggers Erk nuclear translocation, resulting in cell proliferation and differentiation mediated by Erk targets (Robinson et al. 1998; Brunet et al. 1999; Costa et al. 2006). However, it is unclear if a similar translocation occurs following FGF stimulation and MAPK pathway activation in mouse ES cells. Indeed, the effect of FGF signalling in mouse ES cells beyond Mek and Erk, and how this impacts the pluripotency regulatory network, is only just beginning to be elucidated. Erk may function mainly to target and regulate core pluripotency regulatory elements via direct phosphorylation, such as Klf4 (Kim et al. 2012), or also indirectly regulate key transcriptional or protein regulatory mechanisms via additional factors. In addition, visualising Erk localisation and phosphorylation dynamics in the early embryo may also determine whether localisation correlates with, or perhaps even predicts, PrE cell fate specification. Consequently, a principle aim of this project was to characterise FGF/Erk signalling dynamics and their relationship to lineage segregation *in vivo* and ES cell differentiation *in vitro*.

1.1.1 Erk2 phosphorylation correlates with nuclear localisation *in vitro*

To investigate Erk2 localisation patterns in mouse ES cells, Erk2 fluorescent reporter lines were generated (in collaboration with Russell Ernst). Wild type mouse ES cells were transfected with a targeting construct that would fuse a Venus tag in frame with the final C-terminal exon of one Erk allele via homologous recombination (Figure 1.1A). Successfully targeted ES cell colonies

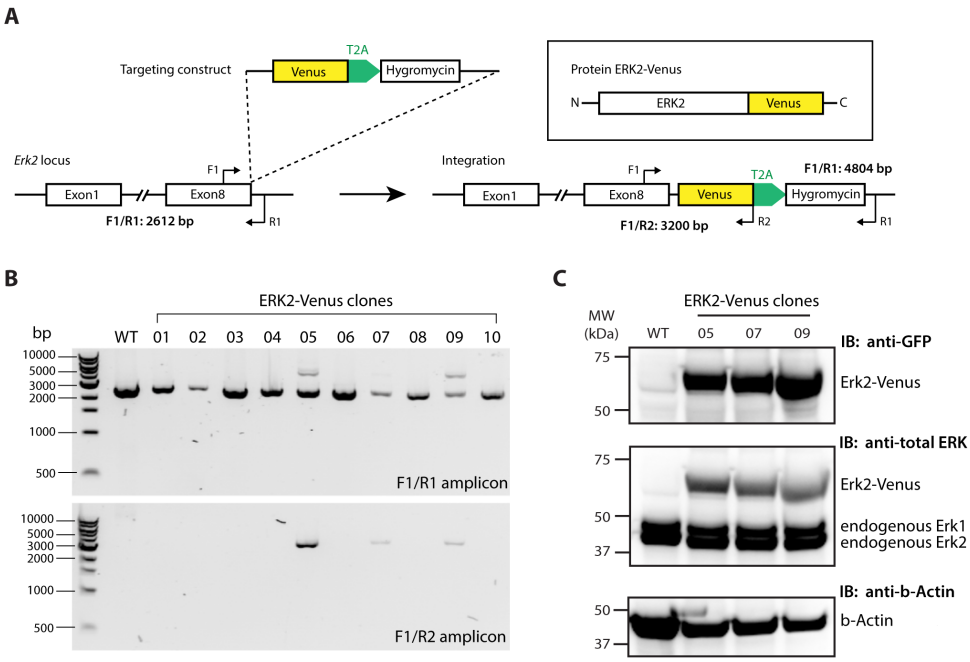


Figure 1.1. Generation of Erk2 reporter mouse ES cell lines. (A) *Erk2::Venus* targeting strategy. The *Erk2::Venus* construct was integrated into Exon 8 of one of the endogenous loci via homologous recombination, generating a heterozygous *Erk2/Erk2*^{Venus} line expressing an ERK2-Venus C-terminal fusion protein. A hygromycin resistance gene is included to enable clonal selection, and separated from the *Venus* sequence by a T2A self-cleaving peptide. F1/R1 and F1/R2 primer pairs were designed to distinguish between wild-type and correctly targeted clones. **(B)** PCR genotyping of wild type (WT) and clonal lines following picking of hygromycin resitant colonies. Results for 3 successfully targeted clones are shown (05, 07, 09). **(C)** Western blot of wild-type ES cells (WT) and successfully targeted *Erk2::Venus* genotyped mouse ESC lines. An Actin loading control is included as a reference.

were selected using hygromycin and correct insertion confirmed by PCR genotyping (Figure 1.1B). Western blot analysis confirmed expression of the Erk2-Venus fusion protein, detected by both anti-GFP and anti-total Erk antibodies (Figure 1.1C).

Previous studies have shown by western blot analysis that Erk phosphorylation in mouse ES cells increases from as early as 10 minutes following FGF stimulation (Hamilton et al. 2013), though it is unclear whether this correlates with nuclear localisation in individual cells. Consequently, the Erk2-Venus fusion presents a promising means of linking localisation dynamics to protein activity. In, preliminary live imaging experiments, Erk2-Venus mouse ES cells were treated with 1 μg/ml FGF2 and examined at defined timepoints to monitor localisation dynamics (Figure 1.2A). Although Fgf2 is not expressed in the early mouse embryo (Guo et al. 2010), both it and Fgf4 function via the Fgfr2 receptor, and these ligands have been used interchangeably to stimulate FGF signalling (Kunath et al. 2005; Cho et al. 2012).

Erk2-Venus was initially localised in the cytoplasm of the majority of cells maintained in serum and LIF, but nuclear translocation was observed following FGF stimulation over a 12-minute period (Figure 1.2A). Immunofluorescence analysis for phosphorylated Erk was next performed to compare untreated cells and cells stimulated with FGF for a longer period. Some phosphorylated Erk was detected in the cytoplasm of untreated cells (Figure 1.2B), likely due to the presence of serum and LIF, which are both known to stimulate Erk phosphorylation (Smith et al. 2004; Kunath et al. 2007). However, nuclear-localised phosphorylated Erk, overlapping with Erk2-Venus, was observed in a number of cells after 30 minutes of FGF treatment (Figure 1.2B). Consequently, FGF stimulation appears to specifically correlate with Erk nuclear localisation, to contrast to serum and LIF.

As FGF stimulation seemingly correlated with Erk2 nuclear localisation *in vitro*, experiments were set up to determine the Erk2 localisation patterns in the E3.5 mouse blastocyst. Given that

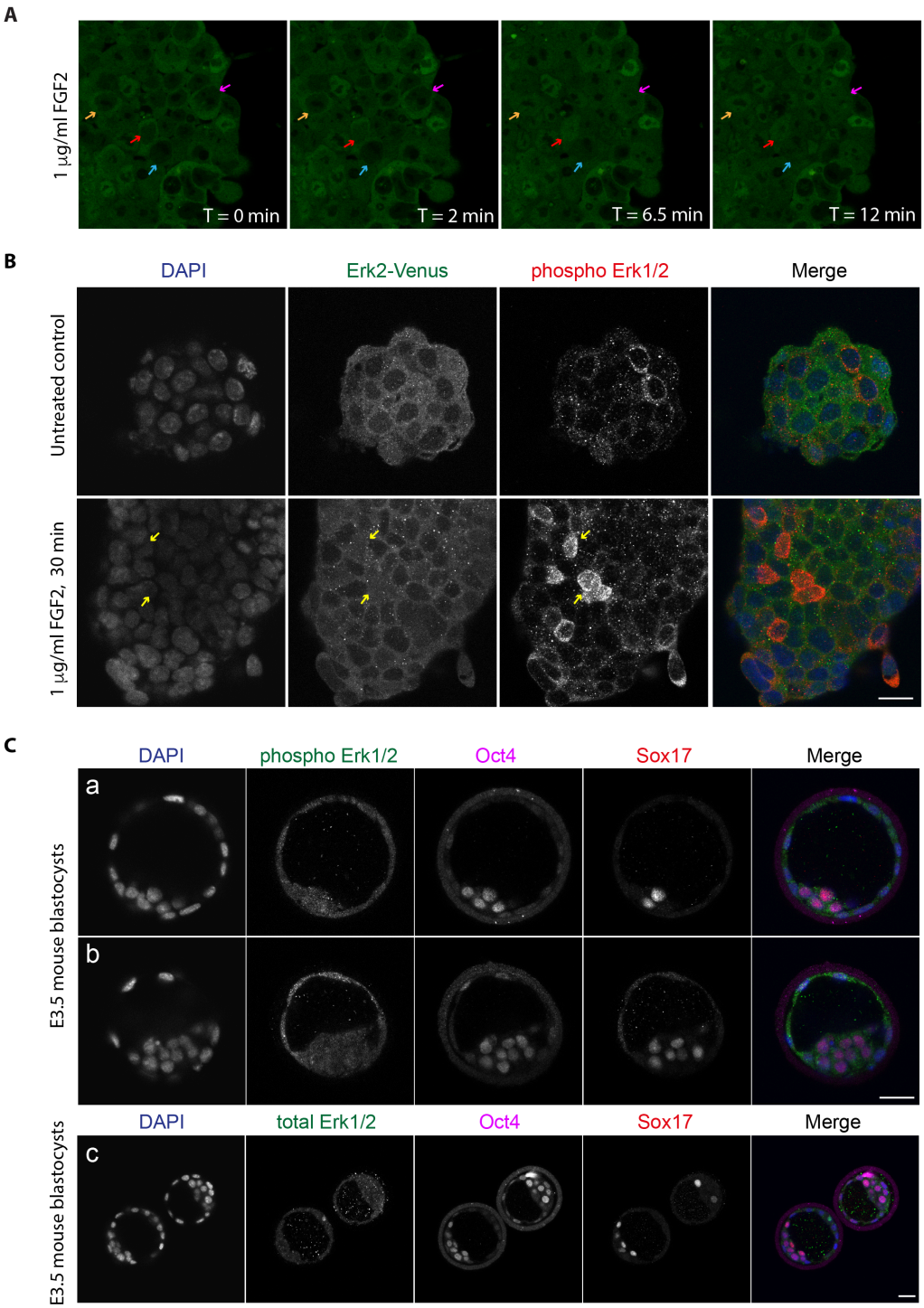


Figure 1.2. Erk2 phosphorylation and localisation patterns. (**A**) Erk2-Venus (green) localisation following stimulation with 1μg/ml FGF2. A single plane is shown from Z-stack images taken at the time points indicated above. Arrows indicate selected examples of cells in which dynamic localisation was observed. n = 3 replicate mES cell lines. (**B**) Phosphorylated Erk (red) in Erk2-Venus (green) mouse ES cell lines with DAPI merge (blue) in serum and LIF (top panel) or stimulated with 1 μg/ml FGF2 for 30 minutes, with arrows indicating examples of dual nuclear localisation. n = 3 replicate mES cell lines. (**C**) Representative Oct4 (magenta), Sox17 (red) and phosphorylated Erk (green; a,b) or total Erk (green; c) with DAPI merge (blue) in E.35 mouse embryos. 5 (total Erk) or 10 embryos (phosphorylated Erk) were analysed.

stimulating or inhibiting FGF signalling in vivo results in conversion of ICM cells to either all PrE or Epi cells respectively (Chazaud et al. 2006; Nichols et al. 2009; Yamanaka et al. 2010), Erk2 nuclear localisation may correlate with cell fate in the embryo. E3.5 blastocysts were stained for phosphorylated or total Erk, and Sox17 and Oct4 were used to identify PrE and Epi cells respectively within the ICM (Figure 1.2C). Oct4 is expressed in all ICM cells at this developmental stage (Dietrich and Hiiragi 2007), so Oct4-positive Sox17-negative cells within the ICM were designated as Epi cells. Although Nanog is a more specific epiblast marker, antibody species restrictions prevented it from being used in tandem with phosphorylated Erk.

Curiously, phosphorylated Erk was not specifically nuclear localised in either Epi or PrE cells (Figure 1.2C), and all cells within the embryo exhibited dispersed phosphorylated Erk distribution. The same distribution pattern was observed for total Erk protein (Figure 1.2C). This may indicate that although a latent level of phosphorylated Erk is present in all cells in the embryo, alternative factors influence an Epi or PrE cell fate choice, which is then reinforced by Erk signalling. Alternatively, differential Erk localisation could be present earlier than the E3.5 stage that was analysed, or could be dynamically oscillating such that the finer details would not be captured by static timepoints.

To better investigate the dynamics of Erk2 localisation during early embryonic development, it was decided to establish an Erk2-Venus reporter mouse line by introducing Erk2-Venus ES cells into recipient embryos to generate chimeras and then subsequent breeding to derive a homozygous line. Live imaging of Erk-Venus reporter embryos would elucidate Erk localisation patterns during early embryonic development, and enable investigations into how modulating FGF signalling alters any observed translocation dynamics. Furthermore, $Erk2^{Venus/Venus}$ mice could also be crossed with Epi or PrE reporter lines to tally localisation with cell fate specification. This work is still on going.

1.2. FGF signalling and lineage specifiers in mouse ES cells

Whilst the initial Erk localisation experiments were in progress, it was reported that Erk signalling *in vitro* was only specifically required for endodermal differentiation, where its primary function was to suppress pluripotency (Hamilton and Brickman 2014). *Nanog* downregulation following Erk signalling activation preceded any detectable increase in the expression of the PrE-associated factor *Gata6*. This is consistent with studies suggesting that Erk may indirectly promote upregulation of an ExEn gene network by suppressing pluripotency factors such as Nanog (Hamazaki et al. 2006), which has been shown to repress Gata6 (Singh et al. 2007). Consistent with this, previous studies demonstrated that Erk directly phosphorylated and inhibited the transcriptional activity of the pluripotency-associated factor Klf4 (Kim et al. 2012), which has been shown to positively regulate Nanog expression in mouse ES cells (Zhang et al. 2010).

Interestingly, overexpression of Gata6, or the closely related zinc finger transcription factor Gata4, is sufficient to reprogram mouse ES cells into XEN-like cells that contribute to PrE-derived lineages in vivo (Fujikura et al. 2002; Shimosato et al. 2007). Gata6 has also been shown to directly regulate endoderm-associated genes such as Hnf4a and Dab2 (Morrisey et al. 1998; Morrisey et al. 2000). In addition, XEN cells can be generated by growth factor-mediated conversion protocols that require FGF signalling, and $Gata6^{-/-}$ cells are refractory to this conversion (Cho et al. 2012; Niakan and Eggan 2013). Reprogramming mouse ES cells to XEN cells involves downregulation of the existing pluripotency network (Fujikura et al. 2002; Shimosato et al. 2007; Cho et al. 2012), though how this occurs remains unclear, and it is possible that Erk and Gata6 function in tandem to drive this fate switch. To explore this further, I sought to next investigate the relationship between Gata6 and FGF/Erk signalling during mouse ES cell differentiation.

1.2.1: Gata6 OE can bypass a requirement for FGF signalling

Mouse ES cell lines expressing a single copy of tetracycline/doxycycline-inducible *Gata6* transgene had previously been generated within our lab (Lily Cho, Kathy Niakan) using a site-specific recombination-based integration strategy (Hochedlinger et al. 2005; Beard et al. 2006). Briefly, the *Gata6* cDNA sequence was amplified from eXEN cell lines and integrated into the *Col1a* locus under the control of a tetracycline operator minimal promoter in KH2 mouse ES cells, which had been modified to include an M2-rtTA tetracycline-responsive transactivator under the control of the *Rosa26* promoter. A FLAG-tag was also integrated in order to distinguish the transgene from the endogenous protein. Control mouse ES cells were generated by integrating a gene encoding a red fluorescent protein, *dsRed* into the *Col1a* locus (Wamaitha et al. 2015). The *Gata6* transgene was also integrated (Lily Cho) into *Fgf4*$^{-/-}$ mouse ES cells (Cho et al. 2012; Kang et al. 2013) in a similar fashion. Following doxycycline induction, robust red fluorescence was observed in the control *dsRed* transgenic line, confirming the fidelity of the transgene system (Fig. 1.3A).

For initial characterisation, the wild type *Gata6* transgenic line was treated with doxycycline for 6 days in the presence of LIF and serum, conditions that would otherwise maintain mouse ES cell self-renewal indefinitely in culture. The *Fgf4*$^{-/-}$ *Gata6*-transgenic cells were first grown for three passages in MEF-free conditions with serum-free media in the presence of an Erk and Gsk3 inhibitor together with LIF (2i+LIF, (Ying et al. 2008)), to eliminate the possibility of signalling from exogenous FGFs. *Gata6* expression was then induced for 6 days in serum-free media.

Wild type *Gata6*-induced cells exhibited a rapid morphological conversion (Figure 1.3B), changing from conventional mouse ES cell colonies to single cells with the dispersed, refractile and stellate morphology characteristic of embryo-derived XEN (eXEN) and growth-factor converted XEN (cXEN) cells (Kunath et al. 2007; Cho et al. 2012). XEN-like cells derived

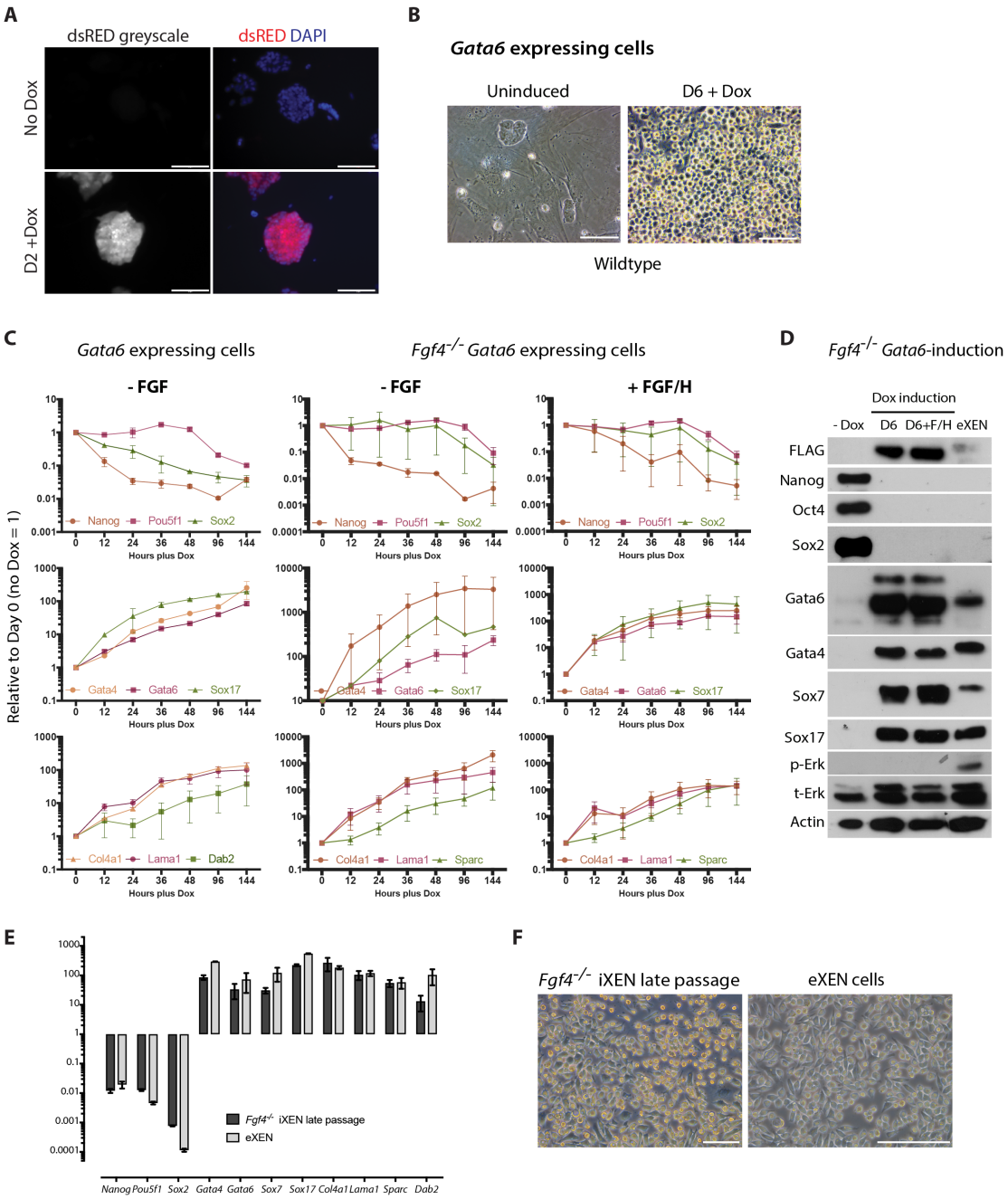


Figure 1.3. Gata6 OE bypasses a requirement for FGF signalling. (A) Immunofluorescence analysis for dsRed (red) with DAPI merge (blue) in *dsRed* transgenic mouse ES cells before and after 2 days of doxycycline treatment. Scale bars: 100 µm **(B)** Representative images of *Gata6* transgenic mouse ES cells before and after 6 days of doxycycline treatment in pluripotency media. Scale bars: 100 µm **(C)** qRT-PCR analysis for selected pluripotency and endoderm transcripts in *Gata6*-induced wild-type and *Fgf4-/-* mouse ES cells between 0 and 144 hours (6 days) of doxycycline induction +/- FGF/H. Relative expression is reflected as fold difference over uninduced mouse ES cells normalized to *Gapdh*. Data are mean ± s.e.m. of 3 biological replicates. (**D**) Western blot for selected proteins in *Gata6*-induced *Fgf4-/-* mouse ES cells +/- FGF/H compared to uninduced mouse ES and eXEN cells. A representative Actin loading control is included. (**E**) qRT-PCR analysis for pluripotency and endoderm transcripts in *Fgf4-/-* iXEN cells maintained for several passages in the absence of doxycycline, compared to eXEN cells. Data are mean ± s.e.m. of 2-3 biological replicates. (**F**) Representative images of stable *Fgf4-/-* iXEN late passage cells and eXEN cells. Scale bars: 100 µm

following transgene overexpression were referred to as induced XEN (iXEN cells) to distinguish them from their embryo-derived counterparts (eXEN cells).

qRT-PCR analysis at defined timepoints over the 6-day timecourse confirmed that both wild type and *Fgf4*$^{-/-}$ *Gata6*-overexpressing cells downregulated pluripotency factors *Nanog, Pou5f1* and *Sox2* over the 6 day timecourse (Figure 1.3C). *Nanog* and *Sox2* transcripts were rapidly downregulated within 12 hours of doxycycline induction, while *Pou5f1* transcript displayed prolonged expression until 48 hours post induction, but was downregulated after 96 hours (Figure 1.3C). Endogenous *Gata6* was also upregulated (distinguished from the FLAG-tagged transgene using primers targeted to the endogenous 3'-UTR sequence), as well as key ExEn genes including *Gata4*, *Sox17*, *Lama1*, *Col4a1* and *Sparc* (Fig 1.3C). This is consistent with observations in *Fgf4*$^{-/-}$ and dominant-negative Fgfr2 mutant mouse ES cells transiently transfected with Gata6 (Li et al. 2004; Kang et al. 2013).

To investigate if exogenous FGF signalling would influence the dynamics of the induction, doxycycline treatment in *Fgf4*$^{-/-}$ *Gata6*-overexpressing cells was repeated in media supplemented with Fgf4 and the FGF receptor binding facilitator heparin (Yayon et al. 1991). Similar gene upregulation and downregulation trends were observed for *Fgf4*$^{-/-}$*Gata6*-induced cells in FGF-supplemented compared to standard media (Fig 1.3C), suggesting FGF signalling has no additive effect. Western blot analysis of selected proteins in *Fgf4*$^{-/-}$ *Gata6*-induced cells after 6 days of induction confirmed loss of pluripotency factors and elevated expression of ExEn genes (Fig 1.3D). Phosphorylated Erk was not detected (Figure 1.3D), suggesting that active Erk signalling is not required for this conversion, and that downregulation of the pluripotency network is occurring via alternative means.

Following the 6-day doxycycline induction, *Fgf4*$^{-/-}$ *Gata6*-induced cells were passaged and switched to XEN cell maintenance media (Kunath et al. 2007) without supplementation with

FGF and heparin, in an attempt to establish a stable iXEN cell line. *Fgf4*$^{-/-}$ iXEN cell lines were maintained for more than two months in the absence of doxycycline without loss of ExEn gene expression or XEN-like morphology (Fig. 1.3E, Fig. 1.3F), suggesting there is no requirement for a feedback loop upregulating Fgf4 to reinforce *Gata6*-mediated reprogramming. Altogether, this suggests that Gata6 can drive a XEN cell fate switch *in vitro* independent of endogenous FGF.

1.2.2. Gata6 OE can also bypass a requirement for Oct4

In addition to its role in promoting pluripotency, Oct4 has also been shown to be required for Fgf4-mediated primitive endoderm specification within the mouse embryo (Frum et al. 2013; Le Bin et al. 2014). Oct4 and Nanog regulate expression of Fgf4 (Nichols et al. 1998; Frankenberg et al. 2011), and although Gata6 expression is initiated in Oct4 or Nanog mutant embryos, the subsequent absence of Sox17 and Gata4 suggests that primitive endoderm formation is compromised (Frankenberg et al. 2011; Frum et al. 2013; Le Bin et al. 2014; Schrode et al. 2014). Although addition of exogenous FGF can restore Sox17 expression in these embryos, later PrE derivatives remain compromised. Previous studies in mouse ES cells showed that Oct4 overexpression upregulated *Gata4* expression (Niwa et al. 2000; Fujikura et al. 2002), although as the mesodermal marker *Brachury* (*T*) was also upregulated (Niwa et al. 2000), this suggests that Oct4 may not necessarily promote solely an endoderm program.

To determine if Oct4 is required to upregulate the wider ExEn network following Gata6 initiation *in vitro*, shRNAs directed against *Pou5f1* were transduced into the wild type *Gata6* transgenic line coincident with doxycycline treatment (Ignacio del Valle). qRT-PCR analysis confirmed significant loss of *Pou5f1* expression compared to scrambled controls (Fig 1.4A), and despite loss of Oct4, *Gata6*-induced cells still upregulated endogenous *Gata6*, *Sox17*, *Gata4*, *Lama1*, *Col4a1*, and *Sparc* and down-regulated pluripotency factors *Sox2* and *Esrrb* (Fig 1.4A).

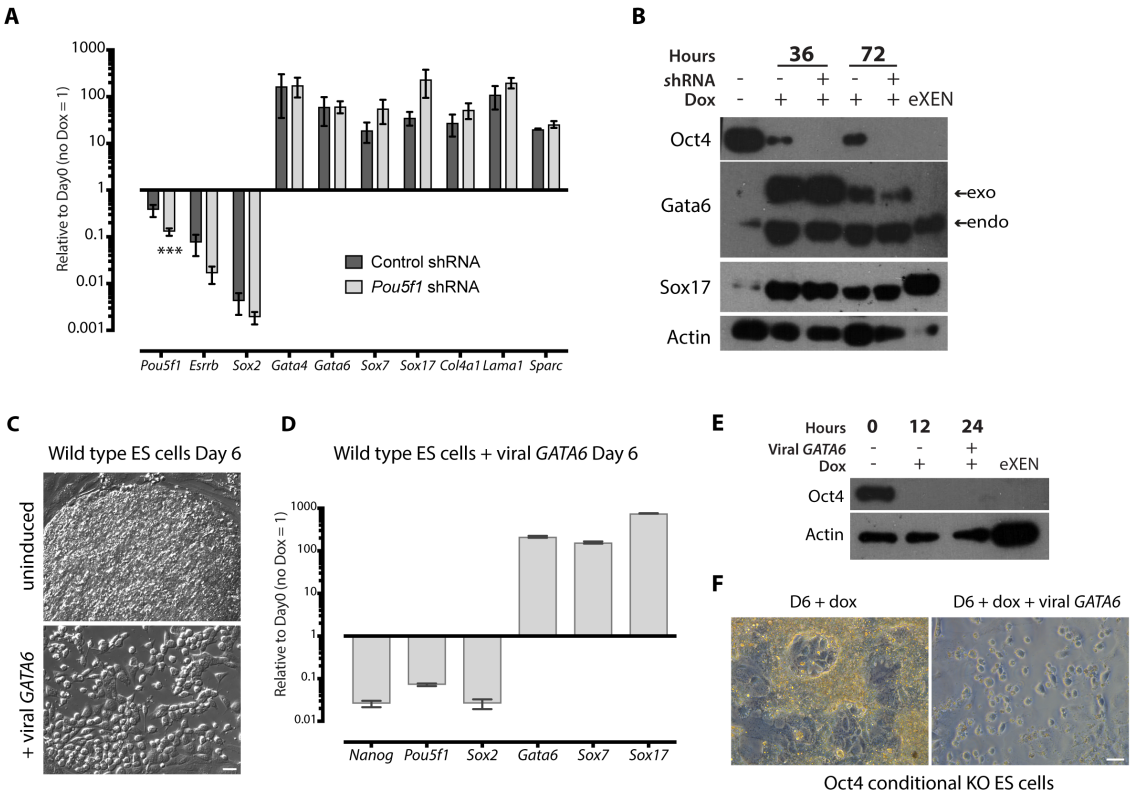


Figure 1.4. Gata6 OE bypasses a requirement for Oct4. (A) qRT-PCR analysis for selected pluripotency and endoderm transcripts following 72 hours of shRNA knockdown of *Pou5f1* during *Gata6* induction compared to scrambled control shRNA. Data are mean ± s.e.m. of 5 distinct shRNA constructs and 2 biological replicates. ***p<0.001. **(B)** Western blot for selected proteins in *Pou5f1* knockdown cells at the time points indicated. A representative Actin loading control is included. **(C)** Representative images of uninduced mouse ES cells compared to 6 days following transduction with a lentivirus driving exogenous expression of human GATA6. **(D)** qRT-PCR analysis for selected pluripotency and endoderm transcripts 6 days following *GATA6* lentiviral transduction in wild type mouse ES cells. Data are mean ± s.e.m. of 2 biological replicates. **(E)** Western blot for Oct4 protein in Oct4 conditional knockout cells at the time points indicated in the presence or absence of exogenous viral HA-tagged *GATA6*. A representative Actin loading control is included as a reference. **(F)** Representative images of Oct4 conditional knockout cells 6 days following doxycycline treatment alone, or in tandem with viral *GATA6* transduction. Scale bars: 20 µm

Western blot analysis verified the expression of Sox17 and endogenous Gata6 protein in the absence of Oct4 in *Pou5f1* shRNA transduced cells, similar to untransduced *Gata6*-induced cells (Fig 1.4B), suggesting Oct4 is dispensable for Gata6-mediated reprogramming.

To further confirm this, an Oct4 conditional knockout mouse ES cell line (ZHBTc4) was identified where doxycycline treatment abrogated Oct4 expression, inducing differentiation to trophectoderm stem (TS) cells (Niwa et al. 2000). To induce Gata6 in ZHBTc4 cells, a lentiviral construct expressing the human *GATA6* gene downstream of an *EF1a* promoter was used, which was shown to successfully reprogram wild type mouse ES cells to iXEN cells (Fig 1.4C, 1.4D), analogous to the doxycycline-inducible system. The *GATA6* construct was transduced coincident with doxycycline treatment and western blot analysis confirmed loss of Oct4 protein within 12 hours of doxycycline addition (Fig 1.4E). Following doxycycline exposure, flattened cells with the cobblestone-like morphology characteristic of TS cells emerged in the untransduced ZHBTc4 cells (Fig 1.4F) as previously observed (Niwa et al. 2000). In contrast, XEN-like cells were observed in the *GATA6*-transduced ZHBTc4 cultures (Fig 1.4F), consistent with the *Pou5f1* knockdown *Gata6*-induced cells. Altogether, this suggests that Gata6 induction can also bypass a requirement for Oct4 in reprogramming mouse ES cell to iXEN cells *in vitro*, and consequently effectively drives this fate switch independently of the mechanisms required for PrE specification in the embryo.

1.2.3. Gata6 (and Gata4) are uniquely capable lineage re-programmers, and exogenous FGF cannot rescue failure of other TFs

To determine whether factors other than Gata6 could also generate iXEN cells, four additional ExEn-associated transcription factors - *Sox7*, *Sox17, Hnf4a* and *Foxa3* - were selected based on their expression in the PrE or its derivatives, and whether previous studies demonstrated a functional requirement for establishing or maintaining the PrE (Chen et al. ; Soudais et al. 1995; Molkentin et al. 1997; Kaestner et al. 1998; Capo-Chichi et al. 2005; Artus et al. 2011; Schrode et

al. 2014). *Gata4* was also included for further characterisation, as it had previously been shown to generate iXEN cells (Fujikura et al. 2002; Shimosato et al. 2007). Transgenic lines were generated (Lily Cho, Kathy Niakan) using the same site-specific recombination-based strategy used for the *Gata6* transgenic lines, and treated with doxycycline for 6 days in the presence of LIF and serum.

Only *Gata4* induction resulted in the emergence of cells with XEN-like morphology, similar to those observed following *Gata6* induction (Fig. 1.5A; Fig. 1.3D), as had been previously observed (Fujikura et al. 2002; Shimosato et al. 2007). Expression of *Sox7*, *Sox17, Hnf4a* or *Foxa3* did not induce a morphological switch during the time period assayed (Fig. 1.5A), and qRT-PCR analysis revealed inconsistent upregulation of ExEn genes, with a failure to upregulate the expression of key factors such as *Col4a1*, *Lama1* or *Hnf4a* to eXEN cell levels (Fig. 1.5B). In contrast, the GATA factors consistently upregulated these genes, including expression of endogenous *Gata6* and *Gata4* (Fig. 1.5B).

Transgene expression was next induced in the presence of exogenous Fgf4 and heparin to test if this could perhaps enhance the likelihood of iXEN conversion, compensating for the inability of other factors to downregulate pluripotency genes. However, again only *Gata4* induction generated iXEN cells within the 6 day time period (Fig. 1.6A). Although the *Sox7*-, *Sox17*-, *Hnf4a*- or *Foxa3*-induced cells lost their mouse ES cell morphology, they seemed to randomly differentiate, and resembled *dsRed*-induced control cells (Fig 1.6A). Additionally, there appeared to be little difference in overall gene expression between FGF-supplemented and LIF-deficient conditions (Fig. 1.6B), where removal of LIF would also destabilize the pluripotency network (Smith et al. 1988; Williams et al. 1988). Only *Gata4*-induced cells generated iXEN cells in LIF-deficient media (Fig 1.6A). Altogether, this indicates that exogenous FGF signalling cannot compensate for the failure of certain endoderm transcription factors to reprogram mouse ES cells to iXEN cells.

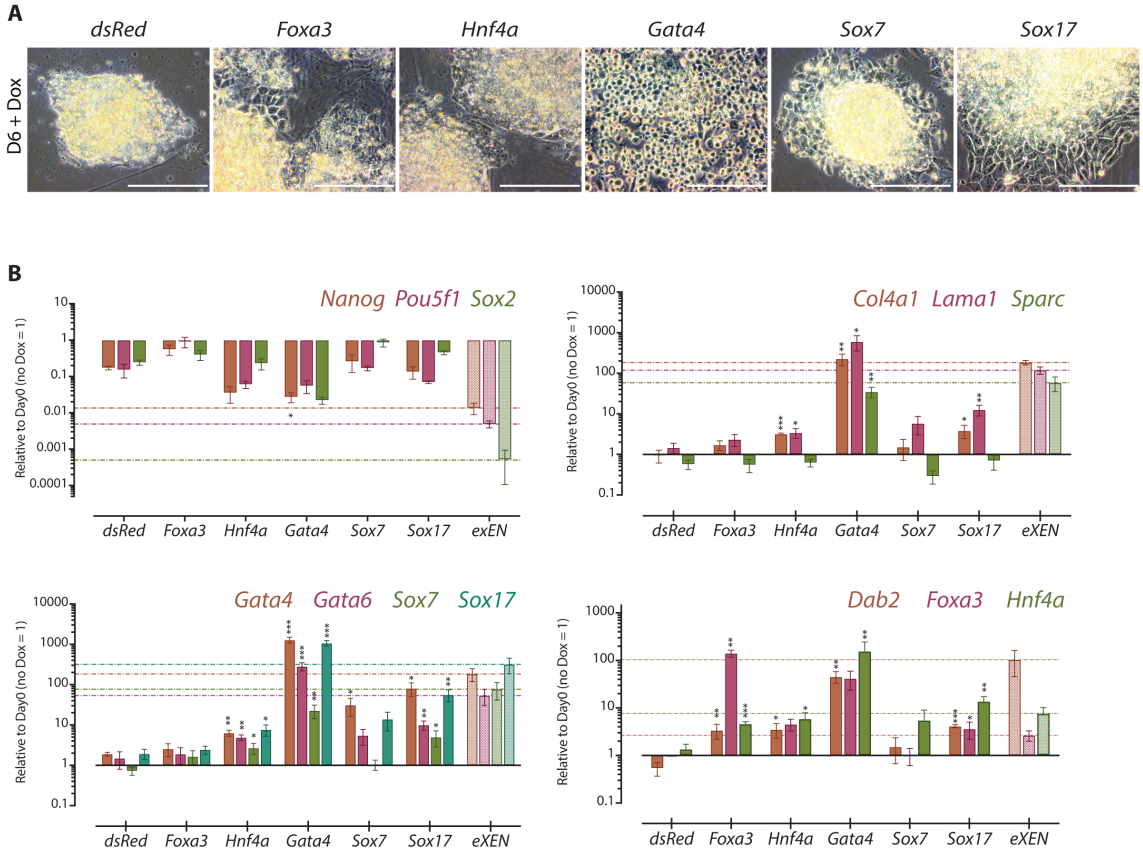


Figure 1.5. Gata6 and Gata4 are uniquely capable lineage reprogrammers. (A) Representative images of *dsRed*, *Foxa3*, *Hnf4a*, *Gata4*, *Sox7* and *Sox17* transgenic mouse ES cells after 6 days of doxycycline treatment in pluripotency medium (serum and LIF). Scale bars: 100 µm. **(B)** qRT-PCR analysis for selected pluripotency and endoderm transcripts in *dsRed*, *Foxa3*, *Hnf4a*, *Gata4*, *Sox7* and *Sox17*-induced cells after 6 days of doxycycline treatment. Relative expression reflected as fold difference over uninduced mES cells normalized to *Gapdh*. eXEN cells are included for comparison, grid lines reflect mean gene expression level in eXEN cells. Data are mean ± s.e.m. of 2-3 biological replicates and 4 technical replicates. *p<0.05, **p<0.01, ***p<0.001. A subset of the qRT-PCR data was collected by Lily Cho and Kathy Niakan.

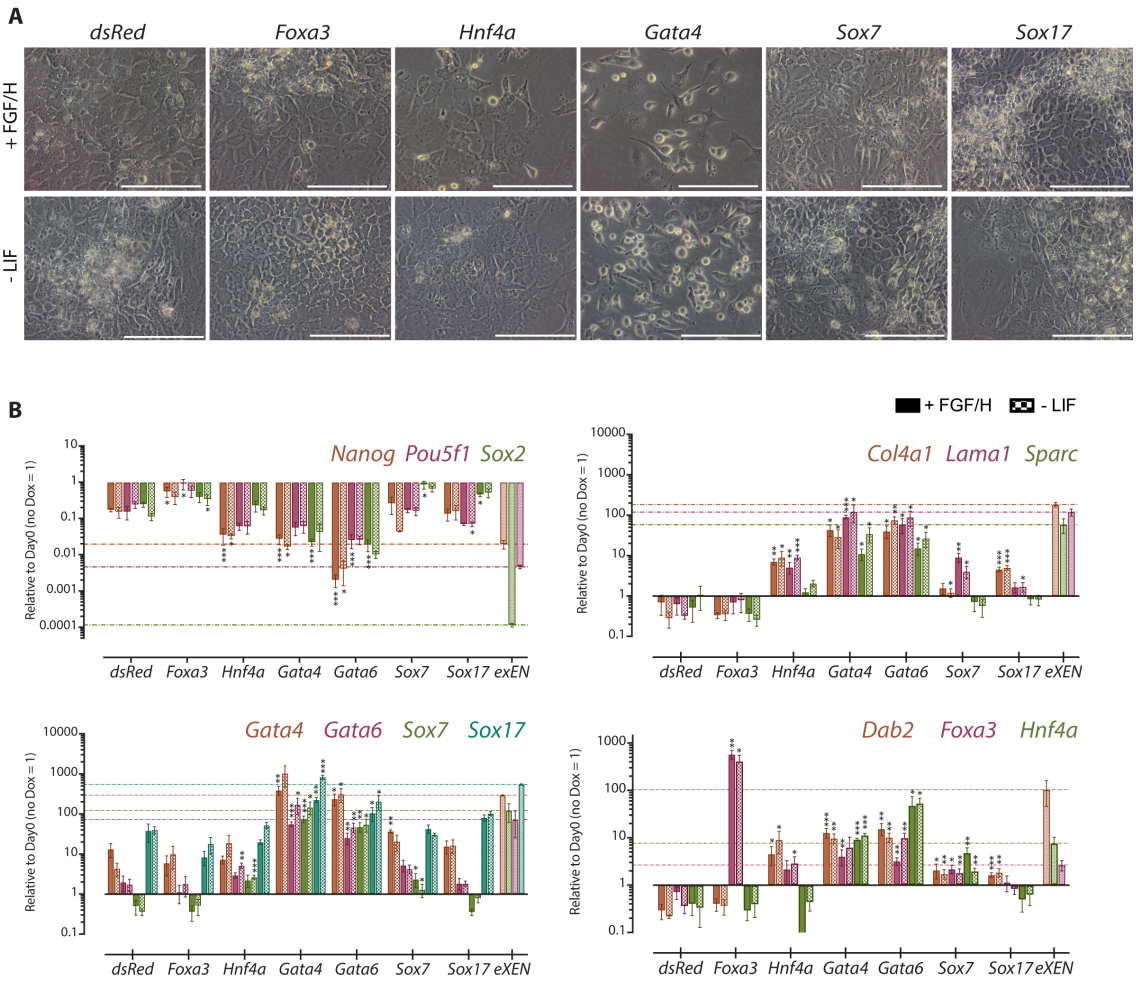


Figure 1.6. Exogenous FGF cannot rescue failure of alternative TFs to drive iXEN reprogramming. (A) Representative images of *dsRed*, *Foxa3*, *Hnf4a*, *Gata4*, *Sox7* and *Sox17* transgenic mouse ES cells after 6 days of doxycycline treatment in FGF/H supplemented or LIF-depleted medium. Scale bars: 100 µm. **(B)** qRT-PCR analysis for selected pluripotency and endoderm transcripts in *dsRed*, *Foxa3*, *Hnf4a*, *Gata4*, *Sox7* and *Sox17*-induced cells after 6 days of doxycycline treatment in FGF/H supplemented or LIF-depleted medium. Relative expression reflected as fold difference over uninduced mES cells normalized to *Gapdh*. eXEN cells are included for comparison, grid lines reflect mean gene expression level in eXEN cells. Data are mean ± s.e.m. of 2-3 biological replicates and 4 technical replicates. *p<0.05, **p<0.01, ***p<0.001. A subset of the qRT-PCR data was collected by Lily Cho and Kathy Niakan.

Previous studies observed that although *Sox17* induction can upregulate ExEn gene expression in mouse ES cells, *Sox17*-overexpressing cells retained mouse ES cell-like morphology and pluripotency factor expression after 48 hours of induction (Niakan et al. 2010; McDonald et al. 2014). Similarly, levels of *Pou5f1, Nanog* and *Sox2* in *Sox17-, Sox7-* or *Foxa3*-induced cells were comparable to the expression in *dsRed* control cells after 6 days of induction (Fig. 1.5B). To confirm this at the protein level, the expression of select pluripotency and ExEn factors was analysed in *Gata6-*, *Sox7-* and *Sox17*-induced cells. Immunofluorescence analysis confirmed that *Gata6*-induced cells expressed Gata4, Gata6, Sox17 and Laminin proteins, and no longer expressed Oct4 (Fig. 1.7A). In contrast, *Sox7*-induced cells persistently expressed Oct4 protein, with only a few cells at the periphery of mouse ES-like colonies expressing Gata4, Gata6, Sox17 and Laminin (Fig. 1.7A). This latter expression pattern was reminiscent of the sporadic differentiation to XEN-like cells previously observed in wild type mouse ES cell cultures (Niakan et al. 2010). Consistent with this, a recent study also demonstrated that *Sox7*-induction is insufficient to generate XEN-like cells (Kinoshita et al. 2015).

Western blot analysis revealed that despite robust induction of exogenous Sox17 over the 6-day treatment period, *Sox17*-induced cells retained expression of Nanog and Oct4 (Fig. 1.7B), and showed delayed upregulation of Gata4 compared to *Gata6*-induced cells (Fig. 1.7B, 1.7C). This is consistent with recent observations that Sox17-expressing cells only acquired XEN-like morphology and downregulated pluripotency factor expression after 12 and 18 days of induction, respectively (McDonald et al. 2014). *Gata6*-induced cells however, downregulated Nanog, Sox2 and Oct4 protein within 12, 36 and 96 hours after induction, respectively (Fig. 1.7C). This strongly suggests that the ability of Gata6 and Gata4 to successfully drive iXEN reprogramming is linked to their capacity to both establish a wider ExEn gene network and downregulate the existing pluripotency program within a shorter time frame.

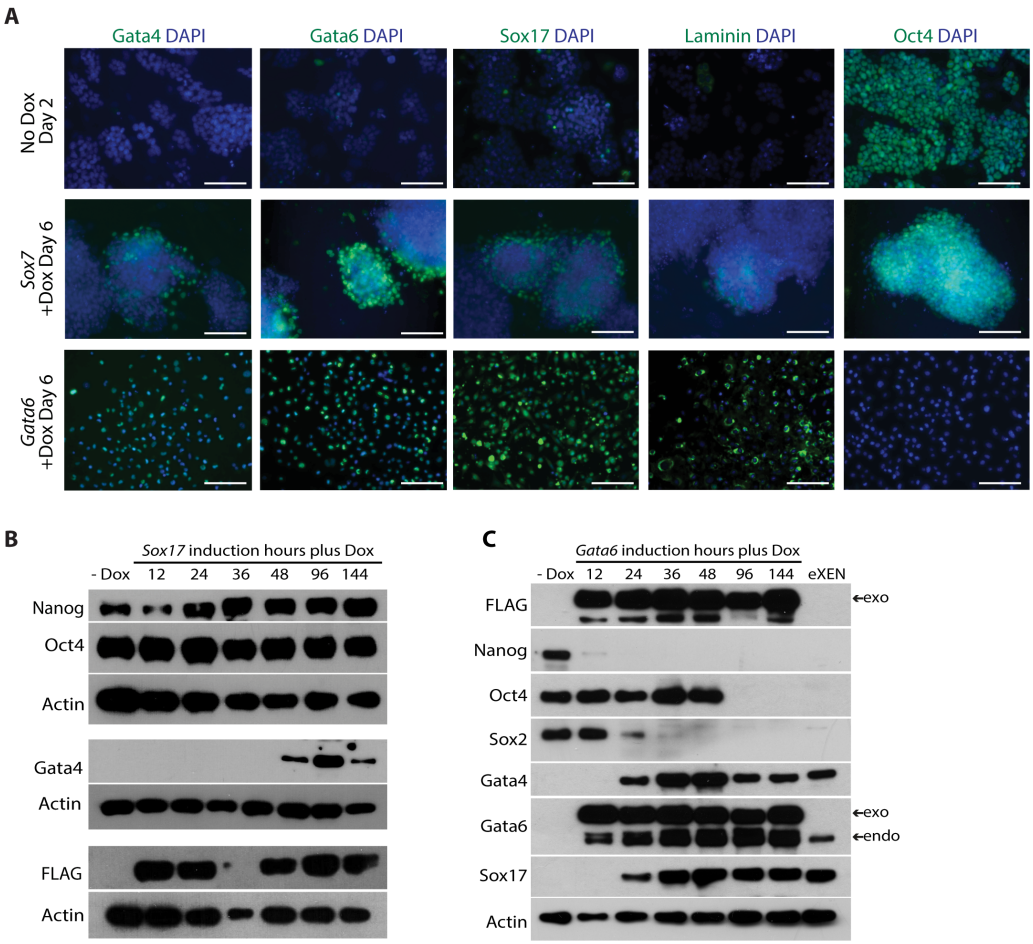


Figure 1.7. *Sox7*- and *Sox17*-induced mouse ES cells fail to downregulate pluripotency factors. (A) Immunofluorescence analysis for Gata4, Gata6, Sox17, Laminin or Oct4 (all green) with DAPI merge (blue) in uninduced mES cells, Sox7 or Gata6 overexpressing mES cells after 6 days of doxycycline treatment. Scale bars: 100 µm. Images were taken by Lily Cho. **(B)** Western blot for selected proteins in *Sox17* transgenic mouse ES cells from 0 to 144 hours of doxycycline induction. Representative Actin loading controls are included. **(C)** Western blot for selected proteins in *Gata6* transgenic mouse ES cells from 0 to 144 hours of doxycycline induction. A representative Actin loading control is included. Endogenous (endo) and exogenous (exo) Gata6 bands are indicated.

1.3. Characterising Gata6-mediated lineage reprogramming

Despite upregulation of *Gata4* and *Sox17* transcripts within 12 hours of *Gata6* induction (Fig. 1.3C), their proteins were not detected by western blot until 24 hours following doxycycline treatment (Fig. 1.7C). Given that Nanog and Sox2 proteins are downregulated in the absence of detectable Gata4 or Sox17 protein (Fig. 1.7C), this suggests that Gata6 directly mediates initial downregulation of the pluripotency program. Direct regulation could also explain why Gata6 can drive reprogramming independent of FGF/Erk-mediated disruption of the pluripotency network. Consequently, and as Gata6 is thought to lie upstream of Gata4 in the PrE transcriptional hierarchy (Morrisey et al. 1998; Koutsourakis et al. 1999; Chazaud et al. 2006; Plusa et al. 2008; Artus et al. 2011; Schrode et al. 2014), subsequent experiments used *Gata6* transgenic mouse ES cells to characterise how this transcription factor drives an ES to XEN fate switch.

1.3.1. A short Gata6 induction is sufficient to perturb mouse ES cell pluripotency

To unravel the mechanisms of Gata6-mediated reprogramming, initial experiments sought to determine the minimum temporal requirement for Gata6 induction to affect mouse ES cell gene expression. *Gata6*-transgenic cells were stimulated with doxycycline for defined pulses of between 2 and 12 hours, and then fixed for immunohistochemistry analysis immediately following doxycycline withdrawal (Fig. 1.8A). Exogenous FLAG-tagged Gata6 protein expression was detectable by immunofluorescence and western blot analysis 2 to 4 hours following doxycycline addition (Fig. 1.8B, 1.8C).

Although induced cells retained mouse ES cell morphology at these early stages, there was a clear effect on protein expression, with Nanog protein downregulated 8 to 10 hours following induction and Sox17 protein beginning to be upregulated around the same time (Fig. 1.8B, 1.8C). However, the pluripotency marker Esrrb remained detectable (Fig. 1.8B,). Consistent with this, qRT-PCR analysis revealed robust induction of FLAG-tagged *Gata6* 2 hours after

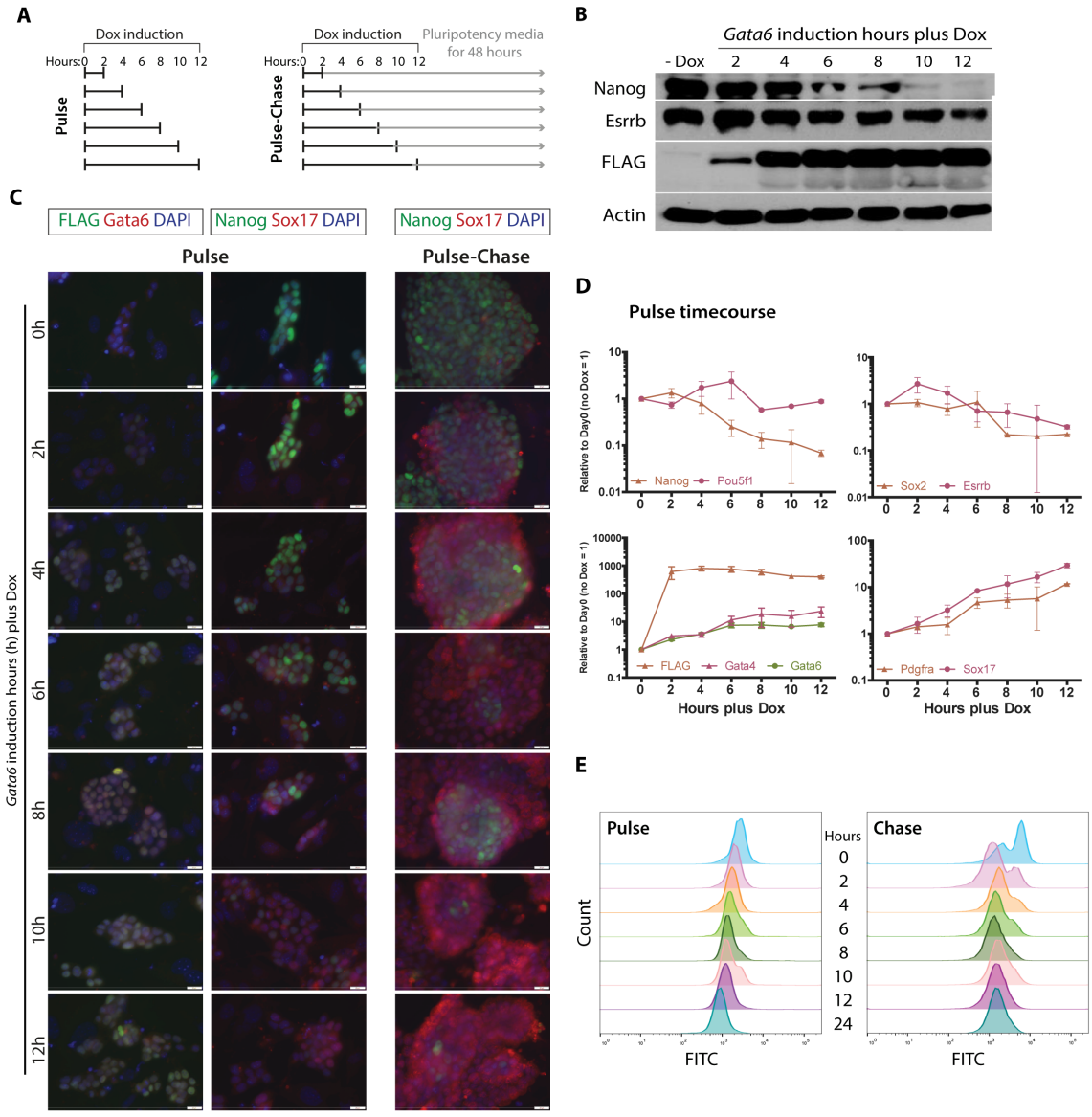
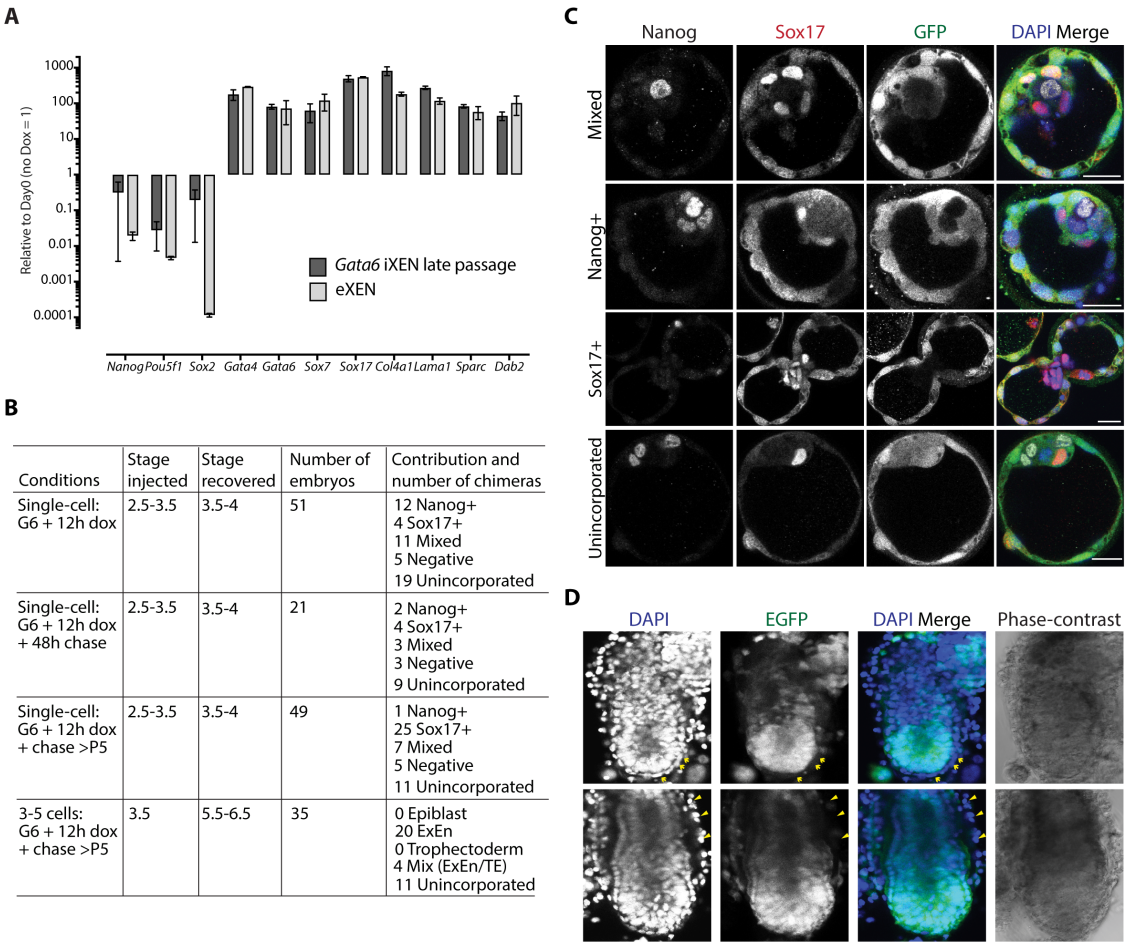


Figure 1.8. A short pulse of *Gata6* induction is sufficient to perturb mouse ES cell pluripotency. (A) Timeline of pulse and pulse-chase induction experiments. Cells were either pulsed for incremental 2 hour time periods and collected for analysis, or switched into pluripotency media for a 48 hour chase period then collected. **(B)** Western blot for selected proteins in *Gata6*-induced cells from 0 to 12 hours of doxycycline induction. A representative Actin loading control is included. **(C)** Immunofluorescence analysis for FLAG (green), Gata6 (red) and DAPI (blue) merge or with Nanog (green), Sox17 (red) and DAPI (blue) merge in *Gata6*-induced cells immediately after doxycycline treatment for defined periods, or following a subsequent 48 hour chase in pluripotency media. Scale bars: 20 μm. **(D)** qRT-PCR analysis for exogenous *FLAG*-tagged *Gata6* expression and selected pluripotency and endoderm transcripts in *Gata6*-induced mouse ES cells between 0 and 12 hours of doxycycline induction. Relative expression is reflected as fold difference over uninduced mouse ES cells normalized to Gapdh. Data are mean ± s.e.m. of 2 biological replicates. **(E)** Flow cytometry analysis of Nanog expression at defined time points in pulse and pulse-chase cells.

doxycycline induction, and greater than 2-fold downregulation of *Nanog* by 6 hours, followed by *Sox2* and *Esrrb* at 8 and 12 hours respectively (Fig 1.8D). *Pou5f1* expression was unchanged at these early time points (Fig. 1.8D). By contrast, endogenous *Gata6* and *Gata4* were upregulated more than 2-fold relative to uninduced controls within 2 hours of doxycycline induction, followed by *Sox17* and *Pdgfra* 4 and 6 hours after induction respectively.

To determine if a short induction pulse was sufficient for lineage commitment, pulse-induced cells were returned to pluripotency media for a 48-hour chase period following doxycycline withdrawal (Fig. 1.8A). Sox17 expression was detected in the majority of pulse-chase cells initially pulsed for 6 hours or longer, suggesting that *Gata6* induction had successfully initiated a stable endogenous ExEn program by this point (Fig. 1.8C). Consistent with this, flow cytometry analysis revealed a gradual reduction of median Nanog expression in pulse-induced cells, correlated with increasing length of Gata6 pulse (Fig. 1.E). Similar analysis in pulse-chase cells revealed a similar shift (Fig. 1.E), with only a small subset of cells in cultures subjected to pulses of 8 hours onwards re-expressing Nanog when returned to pluripotency conditions (Fig. 1.8C, 1.8E). As heterogeneities in Nanog expression are known to exist within mouse ES cell cultures (Chambers et al. 2007; Singh et al. 2007), this likely indicates that these cells have a higher threshold of pluripotency gene expression to overcome. Stable iXEN cell lines were successfully derived from Gata6 expressing cells induced for 6-12 hours and maintained in the absence of doxycycline for more than 10 passages. qRT-PCR analysis confirmed that stable iXEN cells maintained ExEn gene expression at levels comparable to eXEN cells, and did not re-express pluripotency genes (Fig. 1.9A).

To test the developmental potential of short-pulse iXEN cells, unlabelled iXEN cells were injected into B5/EGFP E2.5-3.5 blastocysts that constitutively express enhanced green fluorescent protein (EGFP) from a CMV early enhancer/chicken beta actin (CAG) promoter (Hadjantonakis et al. 1998). The resulting chimera contribution of incorporated cells was



B

Conditions	Stage injected	Stage recovered	Number of embryos	Contribution and number of chimeras
Single-cell: G6 + 12h dox	2.5-3.5	3.5-4	51	12 Nanog+ 4 Sox17+ 11 Mixed 5 Negative 19 Unincorporated
Single-cell: G6 + 12h dox + 48h chase	2.5-3.5	3.5-4	21	2 Nanog+ 4 Sox17+ 3 Mixed 3 Negative 9 Unincorporated
Single-cell: G6 + 12h dox + chase >P5	2.5-3.5	3.5-4	49	1 Nanog+ 25 Sox17+ 7 Mixed 5 Negative 11 Unincorporated
3-5 cells: G6 + 12h dox + chase >P5	3.5	5.5-6.5	35	0 Epiblast 20 ExEn 0 Trophectoderm 4 Mix (ExEn/TE) 11 Unincorporated

Figure 1.9. A longer period of *Gata6* induction is necessary for full commitment to a XEN program. (A) qRT-PCR analysis for selected pluripotency and endoderm transcripts in *Gata6* iXEN cells maintained for several passages in the absence of doxycycline, compared to eXEN cells. Data are mean ± s.e.m. of 2-3 biological replicates. **(B-D)** Chimera contribution of unlabeled *Gata6*-induced cells injected into B5/EGFP embryos that constitutively express EGFP. iXEN cell injections were performed by Crick GEMS facility staff. Chimera counts were compiled by Kathy Niakan. **(B)** Summary of the number of cells injected, stage of injection and embryo recovery, and chimera contribution. Dissections for E5.5 - 6.5 embryo recovery were performed by Kathy Niakan. **(C)** E3.5-4 chimera blastocysts immunofluorescently analysed for the expression of the primitive endoderm marker Sox17 (red), the epiblast marker Nanog (white), EGFP (green) and DAPI (blue) merge. Scale bars: 100 µm. Images taken by Kathy Niakan. **(D)** E5.5-6.5 post-implantation embryos with unlabeled iXEN cell contribution to the visceral (arrow) or parietal (arrow head) endoderm. Representative images of chimeras with phase-contrast, DAPI (blue) nuclear staining and embryo EGFP expression. Images taken by Kathy Niakan.

analysed in E3.5-4 blastocysts or E5.5-6.5 post-implantation embryos (Fig. 1.9B, 1.9C, 1.9D). Injecting Gata6 12-hour pulse cells gave rise to proportionally fewer E3.5-4 chimera blastocysts with exclusive Sox17 expression compared to 12-hour pulse-chase cells (4/32 versus 4/12), and the latter also contributed less frequently to Nanog-expressing cells in the embryo (Fig. 1.9B). This further suggests that while Gata6 is potent and highly efficient in initiating iXEN cell reprogramming, a sufficient interval is required to commit to an ExEn program.

At E3.5-4, injected late passage iXEN cells (greater than P5) contributed predominantly to Sox17-expressing primitive endoderm cells in chimera blastocysts (25/38) (Fig. 1.9B). At E5.5-6.5, late passage iXEN cells contributed to either the extraembryonic visceral or parietal endoderm in the majority of chimera embryos (20/24) (Fig. 1.9B, 1.9D). Epiblast contribution was not detected (Fig. 1.9B), consistent with previous studies using eXEN cells (Kunath et al. 2005). Altogether, this demonstrated that short-pulse Gata6 iXEN cells were capable of successful ExEn contribution in chimera embryos. However, given that a sufficient interval appeared necessary to fully commit to an ExEn program following Gata6 induction, a longer induction timecourse was analysed to examine the effect on global transcriptional dynamics.

1.3.2. Gata6 is enriched in regulatory regions of both up- and downregulated genes

To investigate gene expression dynamics, microarray analysis was performed at defined time points from 12 to 144 hours (6 days), as complete downregulation of Nanog, Oct4 and Sox2 was previously observed by this timepoint (Fig 1.7C). Samples were compared to uninduced mouse ES cells, which reflected the initial pluripotent state, and to eXEN cells, which served as a reference for ExEn gene expression. Additionally, *Sox7* expressing cells at 144 hours post-induction (which were not converted to iXEN by this point, as detailed above) were included to compare their gene expression to Gata6 induced cells. To determine if Gata6 was directly occupying dynamically regulated genes, chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) analysis of Gata6 binding was carried out after 36 hours of

induction, in order to capture both positive and negative gene regulatory dynamics as observed in western blot analyses above (Fig. 1.7C).

K-means clustering was performed (Yingying Wei) on the scaled microarray data to group differentially expressed genes over the time course into 50 clusters (Fig. 1.10A, Appendix Fig. X1; full gene list available in (Wamaitha et al. 2015) Supplemental Table S1). Several functionally significant pluripotency-associated genes were rapidly and persistently downregulated within 12 hours of *Gata6* induction including *Nanog* and *Sox2* as previously observed, and *Nr5a2*, *Klf2* and *Nodal* (cluster 25) (Fig. 1.10A). Additional pluripotency genes including *Esrrb, Dazl*, *Dlk1*, *Ecsit* and *Jarid1b* (cluster 29) were more gradually downregulated over time compared to cluster 25. *Pou5f1, Lefty1, Dppa4* and *Dppa2* (cluster 18) were more persistently expressed before eventual downregulation, suggesting step-wise downregulation of various nodes of the pluripotency gene regulatory network.

Gata6 induction also upregulated several ExEn transcription factors, cell surface proteins and basement membrane components in a step-wise manner (Fig. 1.10A). Initial upregulation of *Gata6*, *Sox17*, *Sox7*, *Foxa1*, *Foxa2* and *Pdgfra* (cluster 15) was followed by upregulation of *Gata4*, *Hnf1b*, *Hnf4a* and key cell surface and basement membrane components *Dab2* and *Lama1* (cluster 39) that are thought to confer an adherence difference to primitive endoderm cells (Gerbe et al. 2008; Niakan et al. 2010; Artus et al. 2011). Additional basement membrane components and ExEn genes including *Lamb2*, *Col4a1*, *Cited1* and *Braf* (cluster 38) were upregulated later in the time course (Fig. 1.10A). Key markers of ectodermal (*Nestin, Pax6, Sox1, Sox3*) or mesodermal (*Flk1, Hand1, Mixl1, Nkx2.5, T*) lineages were not identified within the microarray dataset, suggesting that Gata6 specifically induced an endoderm fate.

By contrast, *Sox7*-induced cells largely retained gene expression patterns similar to mouse ES cells. Although *Sox7*-induced cells upregulated some genes associated with XEN cell function,

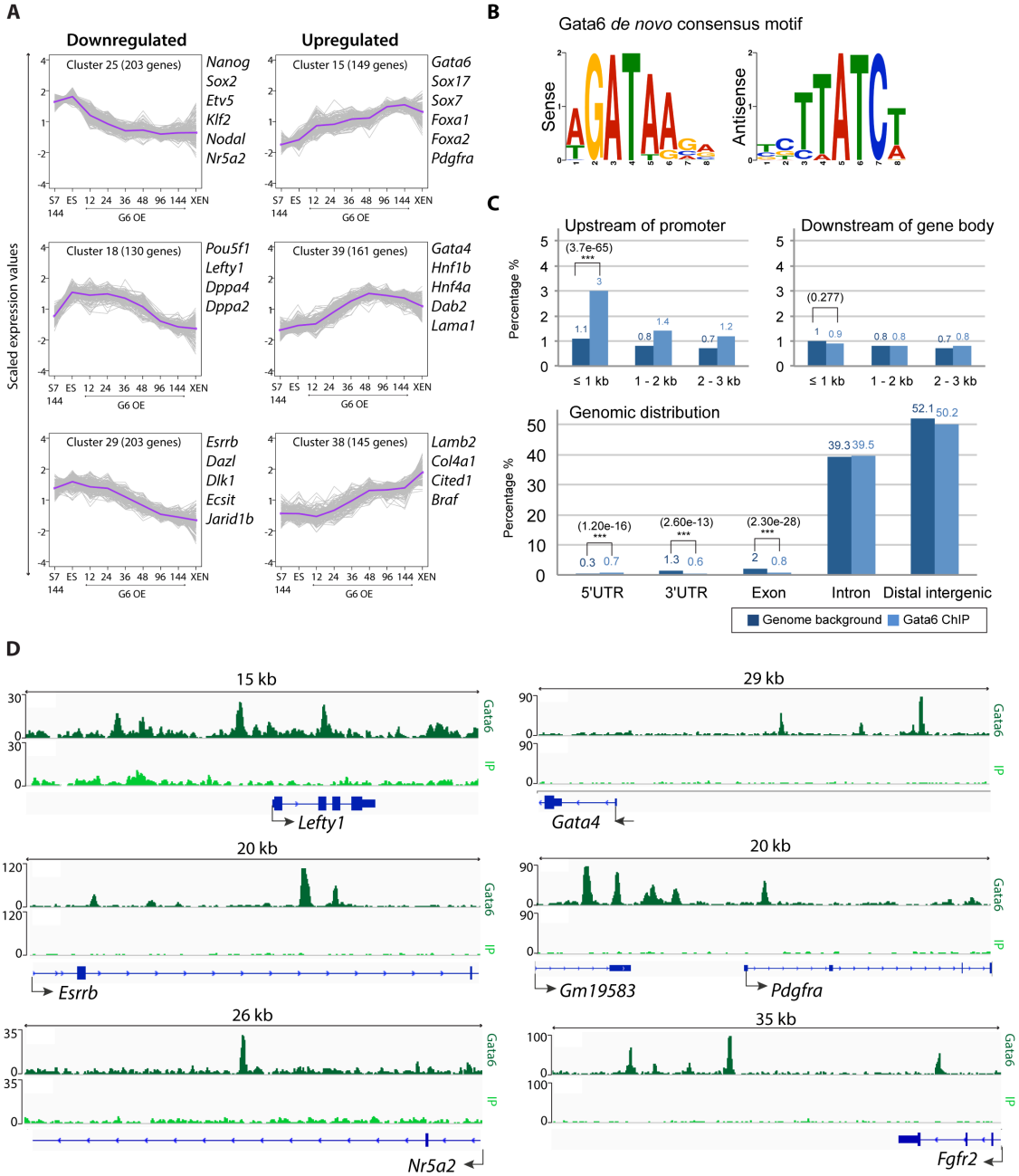


Figure 1.10. Gata6 is enriched in regulatory regions of up- and downregulated genes. (A) Line plots of mean scaled gene expression values from K-means clustering analysis of microarray data from uninduced mouse ES cells, *Gata6*-induced cells from 12 to 144 hours of doxycycline treatment, *Sox7*-induced cells after 144 hours and eXEN cells. Genes were grouped into 50 clusters; selected up- or downregulated clusters containing key pluripotency (left) or endoderm (right) genes are shown here. See Appendix for full dataset. Trajectory of scaled expression values for individual genes are shown as grey lines, while purple lines correspond to the cluster mean. Microarray and K-means analysis performed by Yingying Wei. **(B)** MEME de novo motif analysis of the top Gata6 bound regions following ChIP-seq analysis. MEME analysis performed by Kathy Niakan. **(C)** Comparison of Gata6 binding distribution (light blue) at selected regions relative to the genomic average (dark blue). * p<0.05, ** p<0.01, *** p<0.001. **(D)** ChIP-seq binding profiles showing Gata6 enrichment (dark green) with the input control profile (IP) included for comparison (light green). Representative ChIP-seq binding profile of 3 biological replicates. Two biological replicate samples collected by Lily Cho and Kathy Niakan. ChiP-seq analysis performed by Ignacio del Valle.

such as *Sall4* (cluster 33) (Lim et al. 2008), they maintained expression of pluripotency factors (clusters 18, 25, 29) and did not upregulate endoderm associated genes and basement membrane proteins to the same extent as *Gata6*-induced cells (clusters 15, 38, 39) (Fig. 1.10A). The contrasting gene expression dynamics in *Gata6*- and *Sox7*-induced cells strongly suggested that the potency of Gata6 function might result from binding directly to regulatory regions of both up- and downregulated sets of genes to drive this fate switch.

Analysing the ChIP-seq data (Ignacio del Valle) identified 12,632 Gata6 bound regions enriched over the input control that were common between three biological replicates, with an FDR of <0.01% (full gene list available in (Wamaitha et al. 2015), Supplemental Table S2). *De novo* motif analysis on the top 500 most significant Gata6 bound regions identified the canonical GATA motif as most highly enriched (Fig. 1.10B). Gata6 binding <1000bp upstream of gene promoters was significantly enriched compared to the whole genome (p-value $\leq 3.7 \times 10^{-65}$), in contrast to Gata6 binding downstream of genes, which was not significantly different (p-value ≤ 0.277) (Fig. 1.10C).

Importantly, Gata6 was enriched upstream of genes encoding multiple components of the pluripotency regulatory network, such as *Esrrb*, *Lefty1* and *Nr5a2*, whose expression is downregulated during reprogramming (Fig. 1.10D). Gata6 was also enriched at a number of rapidly upregulated ExEn associated genes such as *Gata4* and *Pdgfra* (Fig. 1.10D), further suggesting that Gata6 directly regulates both pluripotency and ExEn genes. Gata6 was also enriched upstream of *Fgfr2*, suggesting that Gata6 also may be directly regulating FGF signalling.

To relate Gata6 binding to global gene expression dynamics, the ChIP-seq binding data was compared to the microarray cluster dataset (Ignacio del Valle) to identify the subset of Gata6-bound genes that were dynamically regulated over the microarray (full gene list available in (Wamaitha et al. 2015), Supplemental Table S3). The comparative analysis revealed a number of

Gata6-bound rapidly downregulated genes that clustered with known pluripotency factors, suggesting these additional genes may be putative pluripotency factors. One such candidate is *Etv5* (cluster 25), whose expression has also been associated with mouse ES cells (Zhou et al. 2007), but whose function has not yet been tested in this context. However, the absence of Gata6 enrichment near *Sox2*, whose expression is also rapidly downregulated, may suggest indirect repression, possibly due to destabilizing alternative nodes of the pluripotency regulatory network. Alternatively, Gata6 may function as a repressor via a binding site located further away. Nevertheless, Gata6 appears to regulate a wide variety of genes to drive iXEN reprogramming.

1.3.3. Gata6 shares common gene targets and binding sites with pluripotency factors

To determine if Gata6 was competing with pluripotency factors for target gene regulation, spatial heat map analysis (Ignacio del Valle) was used to compare Gata6 bound loci identified in the ChIP-seq analysis to published genome-wide occupancy of Oct4, Sox2, Nanog, Klf4 and Esrrb in mouse ES cells (Fig. 1.11A) (Martello et al. 2013). Venn diagrams also illustrate the overlap between Gata6-bound and pluripotency factor-bound target genes (Fig. 1.11B).

Gata6-bound loci directly overlapped with 136 Oct4, 190 Nanog, 390 Klf4 and 188 Sox2 gene loci (Fig. 1.11A) Intriguingly, some of these sites were present near pluripotency genes including *Lefty1* (Fig. 1.12), suggesting that Gata6 may directly compete with pluripotency factors to antagonize the regulation of some common gene targets. A number of genes, including *Nr5a2* and *Esrrb*, were bound both by Gata6 and one or more of the pluripotency factors, but at discrete loci (Fig. 1.12). Examining the shared gene target dataset ((Wamaitha et al. 2015), Supplemental Table S4) revealed that out of 6192 identified Gata6 target genes, 4195 were also targets of Esrrb, 2815 of Klf4, 1364 of Sox2, 1616 of Nanog and 1104 of Oct4 (Fig. 1.11B), suggesting that Gata6 may regulate the pluripotency network both directly and indirectly.

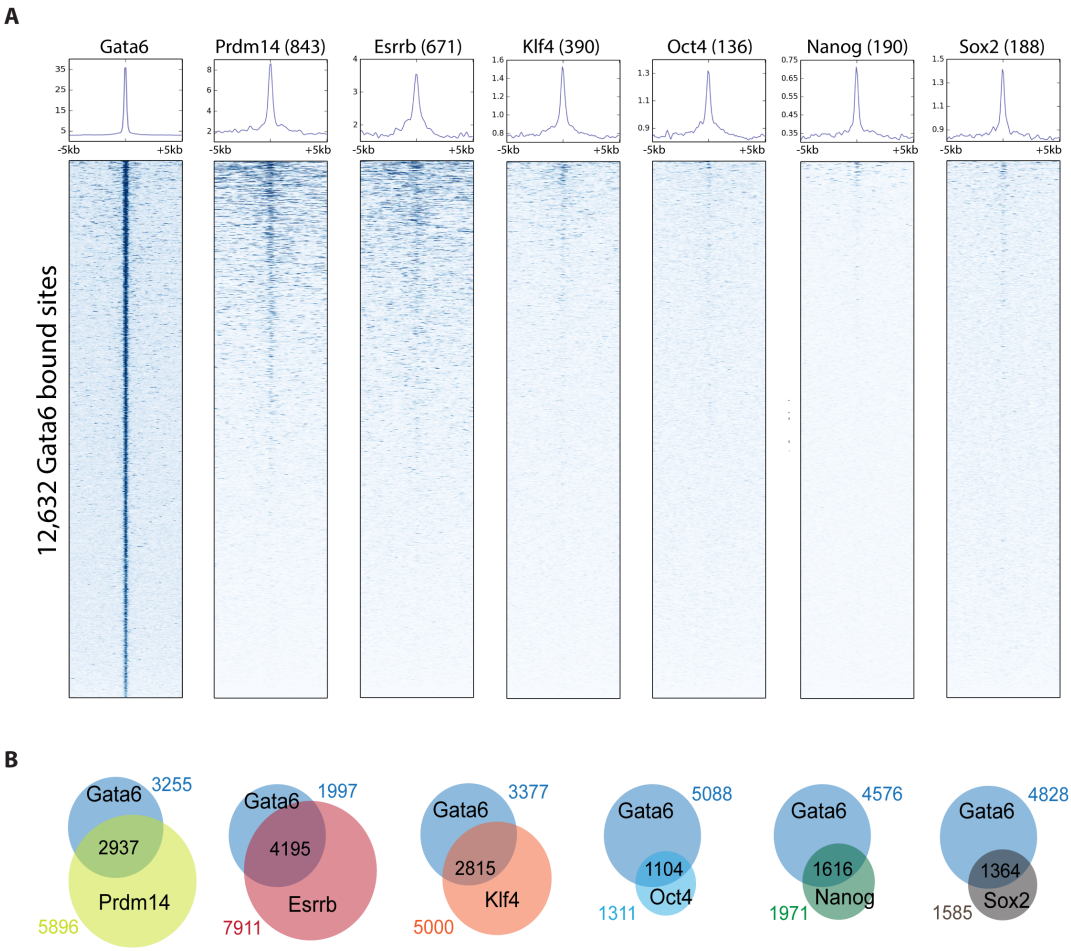


Figure 1.11. Gata6 shares overlapping binding loci and common gene targets with pluripotency factors. (A) Density heat maps of Gata6 binding peak intensity after 36 hours of Gata6 induction indicating direct overlap with Nanog, Oct4, Sox2, Klf4, Esrrb or Prdm14 binding in mouse ES cells within a 10 kb window centered at the transcription start site (TSS). Data for Nanog, Oct4, Sox2, Klf4, Esrrb and Prdm14 was obtained from the Mouse ES Cell ChIP-Seq Compendium (Martello et al., 2012). **(B)** Venn diagram indicating the overlap of Gata6-bound genes during reprogramming compared to genes previously shown to be bound by Oct4, Sox2, Klf4, Esrrb or Prdm14 in mouse ES cells. Spatial heat map analysis and Venn diagrams generated by Ignacio del Valle.

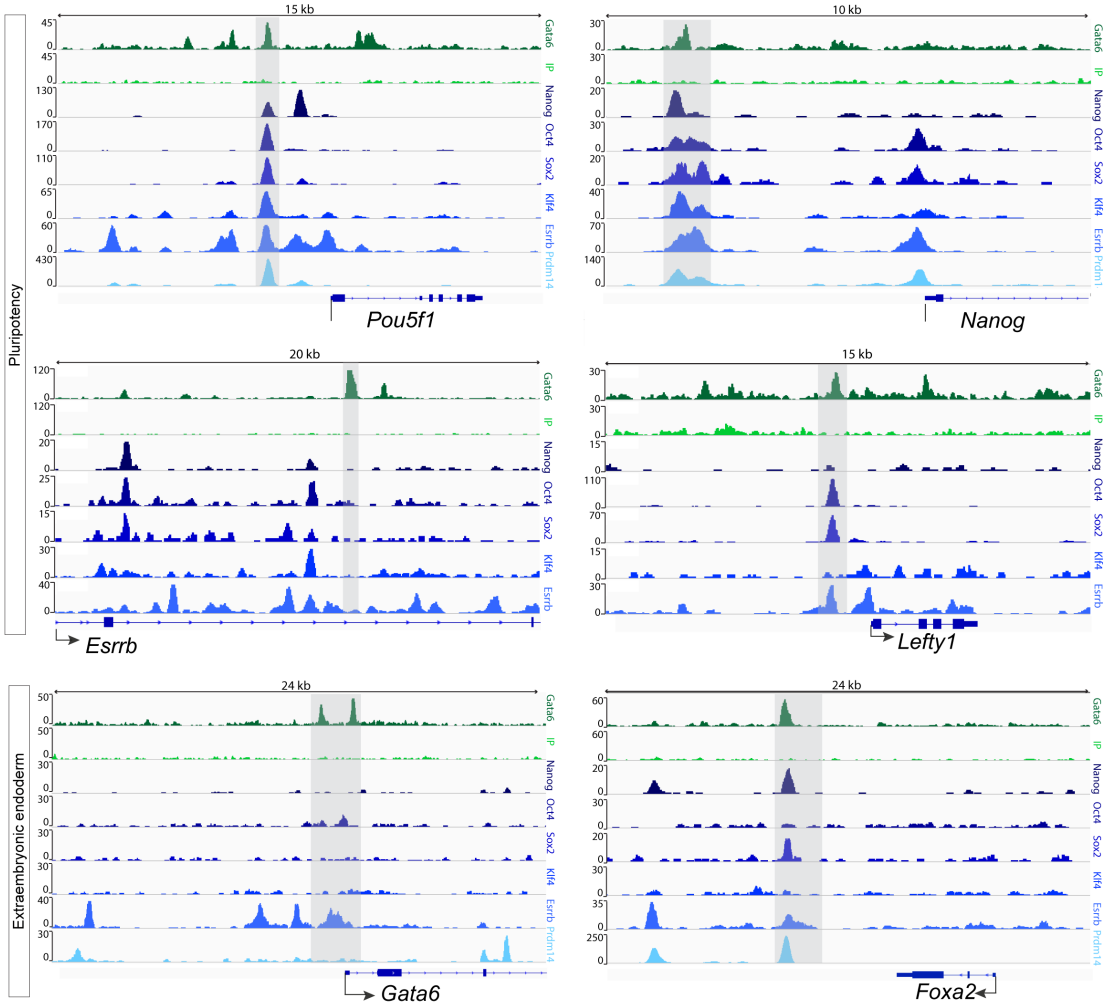


Figure 1.12. Gata6 binding sites overlap with loci previously bound by pluripotency factors in a subset of common gene targets. Binding profiles at various pluripotency and extraembryonic endoderm associsted genes for Gata6 and the input control during reprogramming compared to Nanog, Oct4, Sox2, Klf4, Esrrb or Prdm14 binding profiles in mouse ES cells. Gata6 vs pluripotency factor binding profile track alignment in IGV genome browser performed by Kathy Niakan

Several key ExEn genes were included within the list of Gata6 gene targets shared with Esrrb including *Gata6, Gata4*, *Sox17*, *Col4a1*, *Fgfr2*, *Pdgfra* and *Sox7* ((Wamaitha et al. 2015), Supplemental Table S4). In contrast, most of these key ExEn genes were not identified among the Nanog, Oct4 and Sox2 target gene sets. Previous studies have shown that in addition to its function in pluripotent cells downstream of Nanog, *Esrrb* knockdown or overexpression affects endoderm gene expression (Ivanova et al. 2006; Loh et al. 2006; Festuccia et al. 2012; Uranishi et al. 2013). However, overexpressing an *Esrrb* transgene concomitant with doxycycline-induced *Gata6* expression (Ignacio del Valle) did not prevent downregulation of Nanog expression (Fig. 1.13A) or subsequent iXEN-like cell differentiation. Moreover, immunofluorescence analysis in mouse pre-implantation embryos showed Esrrb and Gata6 expression remained co-incident even after downregulation of Nanog in Gata6-high primitive endoderm cells (Fig. 1.13B).

This suggests that perhaps alternative factors may be required in tandem with Esrrb to block the endoderm promoting effect of Gata6. Knockdown of additional pluripotency factors such as *Prdm14* and *Nr5a2* has also been shown to upregulate ExEn genes (Ma et al. 2011; McDonald et al. 2014) and Gata6-bound loci and gene targets overlapped extensively with those occupied by Prdm14 in mouse ES cells (Fig. 1.11A, 1.11B; ((Wamaitha et al. 2015), Supplemental Table S4), Consequently, the target gene overlap observed may reflect the dual role of Gata6 and a specific pluripotency factor subset in driving their respective cell fates.

Finally, ChIP-seq analysis of Gata6 occupancy in eXEN cells was performed to determine if the binding sites identified during Gata6 reprogramming are maintained in this context. Analysing the ChIP-seq data (Ignacio del Valle) determined that of the 927 Gata6 gene targets in eXEN cells, 504 genes were also targets of Gata6 at the 36-hour post induction timepoint, including endoderm target genes such as *Gata4*, *Gata6*, *Sox7* and *Lamc1* ((Wamaitha et al. 2015), Supplemental Tables S6 and S7). Interestingly, in the eXEN ChIP-seq, Gata6 binding was not detected near many of the pluripotency target genes identified in the induction ChIP-seq, with

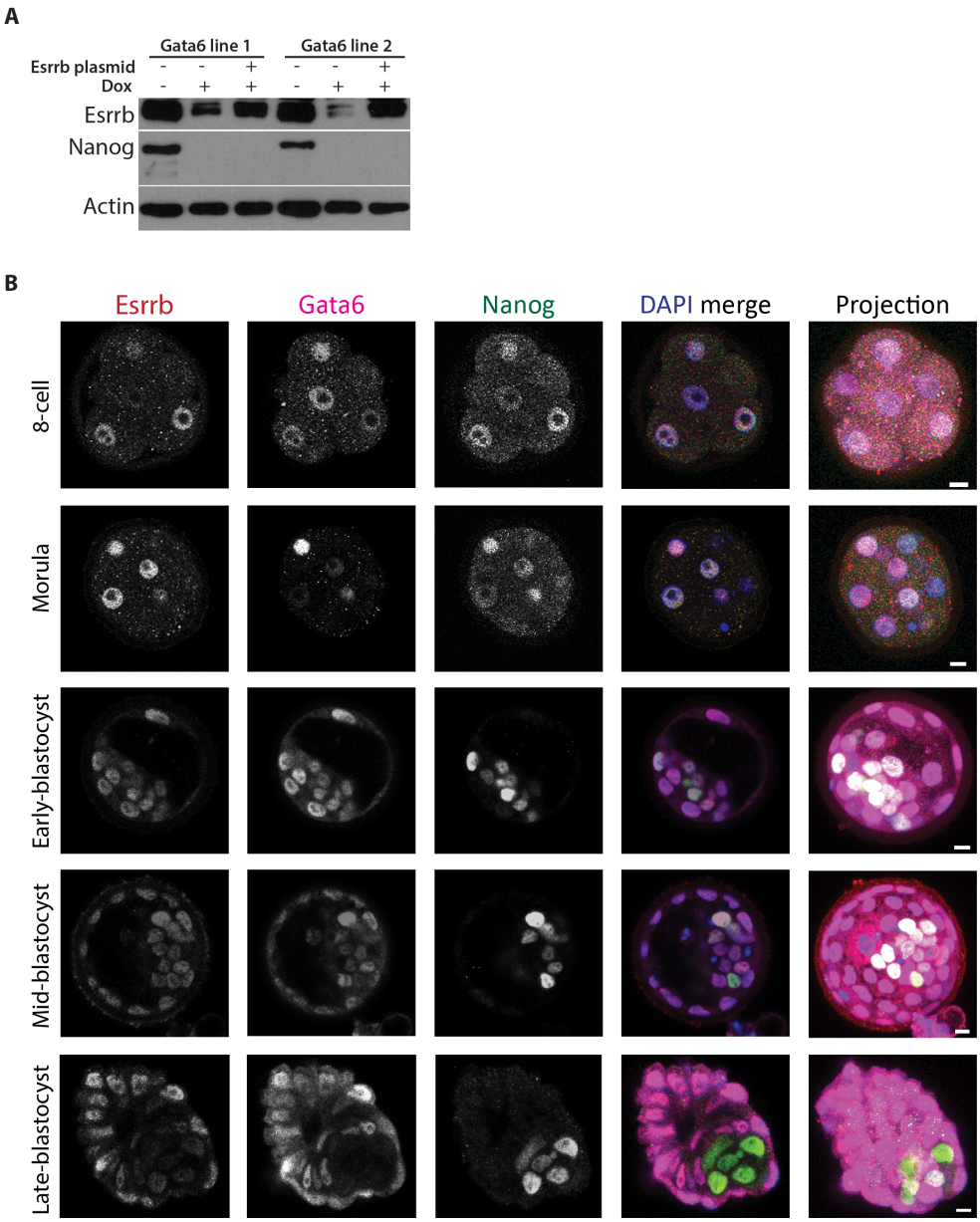


Figure 1.13. *Esrrb* OE in *Gata6*-induced cells does not prevent Nanog downregulation (A) Western blot for selected proteins in *Gata6*-induced or *Esrrb*-transfected *Gata6*-induced mouse ES cells 24 hours after doxycycline treatment, compared to control cells. A representative Actin loading control is included. **(B)** Confocal optical sections and 3D projections of 8-cell, morula, early-, mid- and late-mouse blastocysts. Embryos were immunofluroescently analysed for the expression of Esrrb (red), Gata6 (magenta) and Nanog (green) with DAPI (blue) merge. Scale bar: 10 µm. IHC analysis performed and images in panel B taken by Norah Fogarty. 3 to 5 embryos were analysed per timepoint.

the notable exception of *Nr5a2*. This demonstrates that Gata6 binding in eXEN cells is distinct compared to cells in transition, and suggests there may not be a requirement for Gata6 to actively repress pluripotency factor expression in the long-term after initial downregulation.

1.4. Gata6/GATA6 reprogramming is not restricted to mouse ES cells

ES cells have a characteristic open chromatin state (Chen and Dent 2014), which may make them more amenable to binding of regulatory elements to control gene expression. Consequently, Gata6 may have fewer roadblocks to driving reprogramming in this context, as compared to a more differentiated cell type with a less accessible genomic landscape. To explore the effect of Gata6 induction in a more committed cell type, GATA6 expression was induced in a pure culture of a stable neural stem cell line, NS5 (Conti et al. 2005), using the *GATA6* lentiviral vector previously used to reprogram the ZHBTc4 Oct4 conditional knockout (Fig. 1.4C-F). NS5 cells had been shown to generate functional neurons that contribute to the adult brain in mouse chimeras, without the formation of teratomas (Conti et al. 2005), indicating commitment to their specific lineage. The *GATA6* construct was transduced into NS5 cells in neural basal media and cell identity evaluated 20 days after induction.

Despite retaining a morphological resemblance to neurons (Fig. 1.14A), immunofluorescence analysis confirmed that *GATA6*-transduced NS5 cells robustly induced Gata6, Sox7 and Sox17 proteins (Fig. 1.14B). Neither untransduced nor *GATA6*-transduced cells expressed Oct4 and Nanog (Fig 1.14B, (Conti et al. 2005). Low levels of Nestin remained detectable in *GATA6*-transduced cells (Fig. 1.14B) suggesting that an additional time interval may be required to stabilise the XEN cell program and fully overcome the neural stem cell state. Nevertheless, conservation of GATA6/Gata6-mediated upregulation of at least some gene targets between mouse ES and neural stem cells, despite their distinct supporting gene networks, suggests that Gata6 is a powerful lineage specifier.

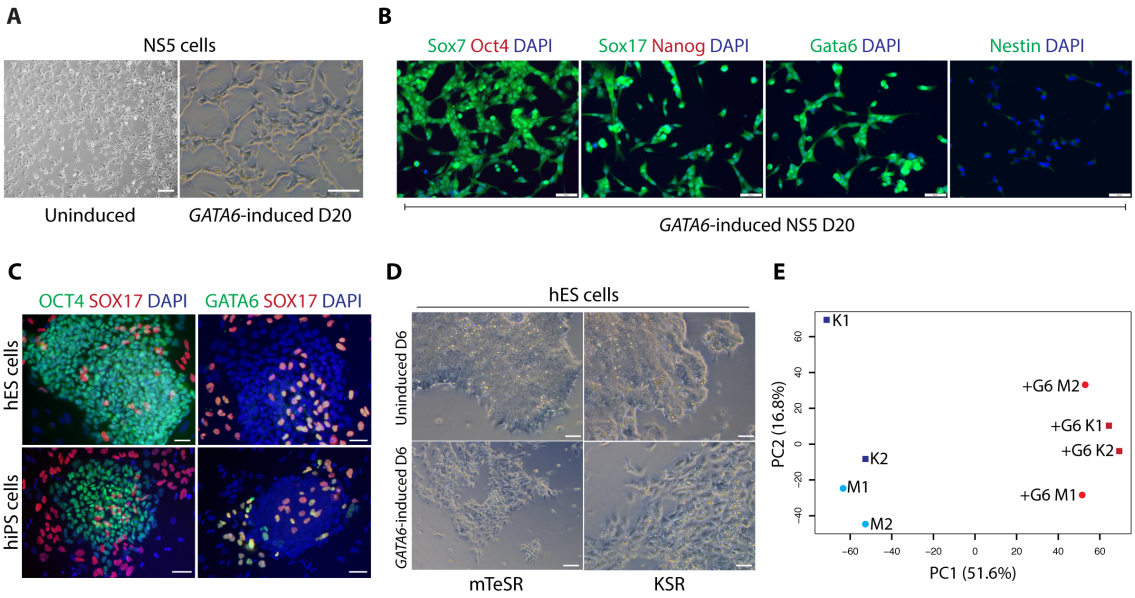
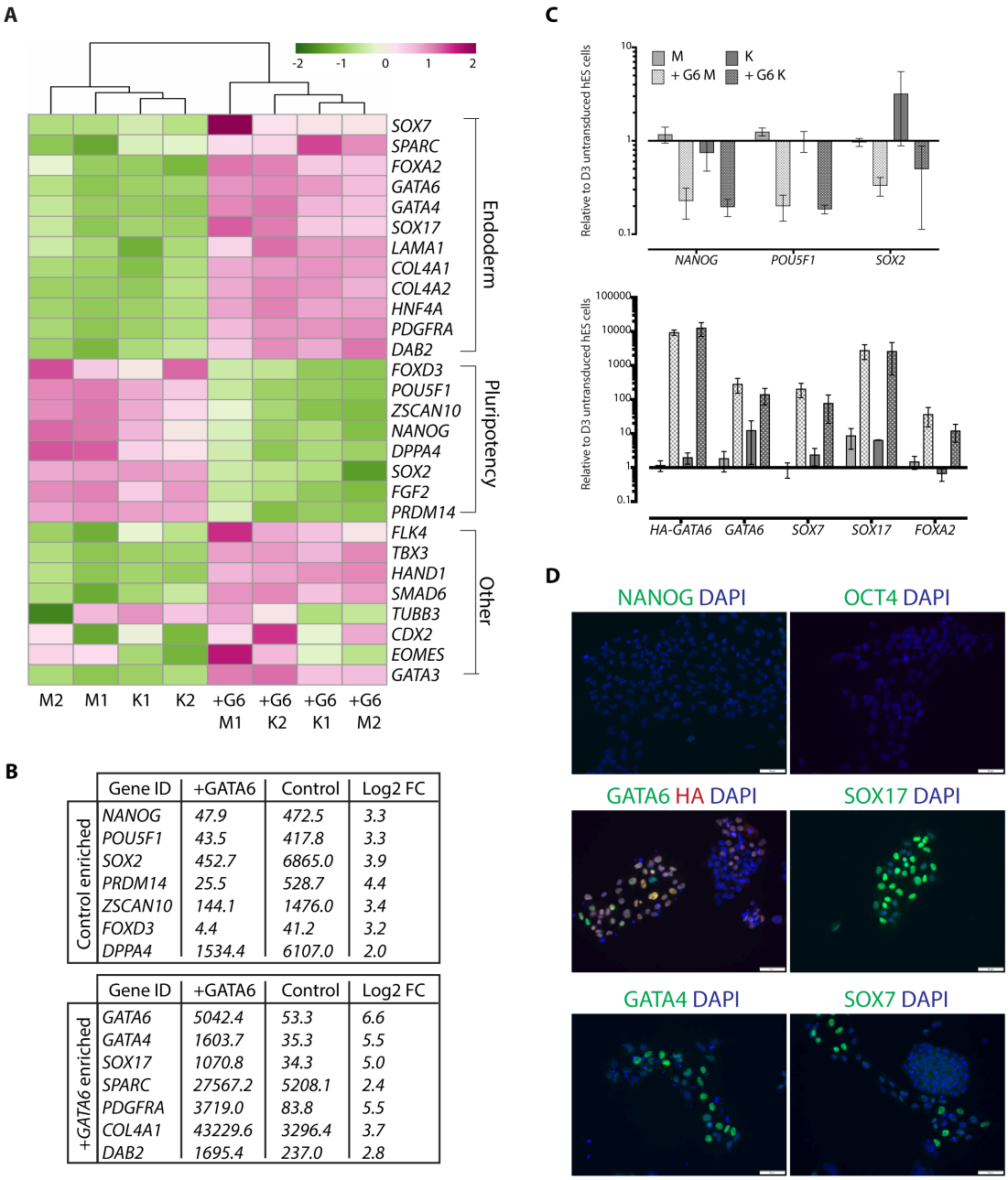


Figure 1.14. Gata6 initiates an ExEn program in mouse neural stem cells and in human ES cells. (A) Phase-contrast images of uninduced or viral *GATA6*-induced mouse NS5 neural stem cells. Scale bars: 50 µm. Uninduced image taken by Isabelle Blomfield. **(B)** Immunofluorescence analysis for Sox7, Sox17, Gata6 or Nestin (green), Oct4 or Nanog (red), with DAPI (blue) merge in *GATA6*-induced mouse neural stem cells after 20 days of induction. Scale bars: 50 µm. **(C)** Immunofluorescence analysis for OCT4 or GATA6 (green) and SOX17 (red) with DAPI (blue) merge in human ES and iPS cells in pluripotent culture conditions. Scale bars: 50 µm. Images taken by Kathy Niakan. **(D)** Representative images of uninduced and *GATA6*-transduced hES cells after 6 days after induction in pluripotency (mTeSR) or permissive (KSR) media. Scale bars: 50 µm. **(E)** PCA using RPKM normalized RNA-seq data from biological replicates of uninduced and *GATA6*-induced (+G) cells in KSR (K) or mTeSR (M) media. PCA performed by Paul Blakeley.

The potency of Gata6 to initiate an ExEn program prompted a final investigation, in a human context. Stable XEN cell lines have yet to be established directly from the human PrE, but SOX17 and GATA6 co-expressing cells can occasionally be observed within human ES and iPS cell cultures maintained in pluripotency conditions (Fig. 1.14C). This is reminiscent of the Sox17 expressing XEN-committed cells observed in mouse ES cell cultures (Niakan et al. 2010) and suggests that human ES cells may also have the potential to be converted to XEN cells, perhaps by ExEn transcription factor overexpression. Although previous studies overexpressing *SOX7* or *SOX17* in human ES cells report upregulation of either an ExEn or embryonic endoderm program respectively, cells retain expression of NANOG and OCT4, and cannot be maintained indefinitely in culture (Seguin et al. 2008). To determine if *GATA6* expression is sufficient to drive iXEN cell reprogramming from human ES cells, a viral *GATA6* construct was transduced into human ES cells in both pluripotency (mTeSR) and differentiation-permissive (KSR) conditions (Ignacio del Valle).

A morphological benchmark for human eXEN cells has yet to be determined. Nevertheless, *GATA6*-transduced human ES cells were morphologically distinct from pluripotent human ES cell colonies in both mTeSR and KSR (Fig. 1.14D), suggesting that these cells are beginning to acquire an alternative identity. RNA-sequencing analysis was performed to compare the global transcriptional profile of *GATA6*-transduced human ES cells to their untransduced counterparts. A principle components analysis (PCA) of genes with the most variable expression revealed that *GATA6*-transduced samples clustered together irrespective of basal media, and were transcriptionally distinct from untransduced cells (Fig. 1.14E, Paul Blakeley). This was confirmed using hierarchical clustering (Fig. 1.15A, Paul Blakeley). Differential gene expression between transduced and untransduced cells was analysed using DESeq (Anders and Huber 2010), an extension of the negative binomial distribution model applied to read count data (Robinson and Smyth 2008; Robinson et al. 2010) (Fig. 1.15B, Paul Blakeley).



B

	Gene ID	+GATA6	Control	Log2 FC
Control enriched	*NANOG*	*47.9*	*472.5*	*3.3*
	POU5F1	*43.5*	*417.8*	*3.3*
	SOX2	*452.7*	*6865.0*	*3.9*
	PRDM14	*25.5*	*528.7*	*4.4*
	ZSCAN10	*144.1*	*1476.0*	*3.4*
	FOXD3	*4.4*	*41.2*	*3.2*
	DPPA4	*1534.4*	*6107.0*	*2.0*

	Gene ID	+GATA6	Control	Log2 FC
+GATA6 enriched	*GATA6*	*5042.4*	*53.3*	*6.6*
	GATA4	*1603.7*	*35.3*	*5.5*
	SOX17	*1070.8*	*34.3*	*5.0*
	SPARC	*27567.2*	*5208.1*	*2.4*
	PDGFRA	*3719.0*	*83.8*	*5.5*
	COL4A1	*43229.6*	*3296.4*	*3.7*
	DAB2	*1695.4*	*237.0*	*2.8*

Figure 1.15. Gata6 initiates an ExEn program in human ES cells. (A) Heatmaps showing hierarchical clustering of uninduced and *GATA6*-induced (+G) cells in KSR (K) or mTeSR (M) media using RPKM normalized RNA-seq data. Expression levels plotted on a high-to-low scale (purple-white-green). Heatmaps were generated by Paul Blakeley. **(B)** DESeq analysis indicating genes significantly differentially expressed in control versus *GATA6*-induced cells in KSR and mTeSR . Median gene expression and log2 fold change (FC) difference in expression is noted. DESeq analysis was carried out by Paul Blakeley. **(C)** qRT-PCR analysis for exogenous *HA*-tagged *GATA6* expression and selected pluripotency and endoderm transcripts 6 days following doxycycline treatment. Data are mean ± s.e.m. of 2-3 replicates from both H1 and H9 human ES cell lines. **(D)** Immunofluorescence analysis of OCT4, NANOG, GATA4, SOX17, SOX17, GATA6 (all green), HA (red), with DAPI (blue) merge in *GATA6*-induced H9 human ES cells 5 days following doxycycline treatment. Scale bars: 50 µm.

Notably, *GATA6*-transduced human ES cells upregulated a number of extra-embryonic and/or pan-endoderm factors including *GATA6*, *GATA4*, *SOX17*, *SOX7*, *FOXA2*, *PDGFRA*, *COL4A1*, *COL4A2*, *LAMA1*, *HNF4a*, *DAB2* and *SPARC* (Fig. 1.15A; 1.15B, (Wamaitha et al. 2015), Supplemental Table S7). Significantly, *GATA6*-transduced cells downregulated the expression of several pluripotency factors including *NANOG*, *POU5F1*, *SOX2*, *PRDM14*, *FGF2*, *FOXD3*, *DPPA4* and *ZSCAN10* (Fig. 1.15A, 1.15B; ((Wamaitha et al. 2015), Supplemental Table S7). However, *GATA6*-transduced cells also expressed genes associated with alternative lineages such as *FLK4*, *TBX3*, *CDX2*, *GATA3*, *SMAD6*, *EOMES*, *HAND1* and *TUBB3* (Fig. 1.15A), suggesting they may not yet be fully reprogrammed to stable iXEN cells. Moreover, although *GATA6*-transduced human ES cells could be clonally passaged more than 3 times and maintained their morphology for more than one month, they could not be maintained indefinitely, suggesting alternative conditions or factors may be required to derive stable human iXEN cells.

Human ES cells were also transduced with an alternative tetracycline/doxycycline-inducible HA-tagged *GATA6* transgene (Ignacio del Valle). Following doxycycline treatment for 6 days, a morphological switch was again observed, and qRT-PCR analysis confirmed the upregulation of *SOX17*, *SOX7*, *FOXA2* and endogenous *GATA6* transcripts, and the downregulation of the expression of *NANOG, SOX2* and *POU5F1* (Fig. 1.15C). Immunofluorescence analysis confirmed downregulation of OCT4 and NANOG, and upregulation of SOX7, SOX17 and GATA4 proteins (Fig. 1.15D), although heterogeneity in endoderm protein induction suggests that an additional time interval, culture condition or factor(s) may be needed to generate self-renewing human iXEN cell lines. Altogether this suggests that GATA6/Gata6 is sufficient to overcome a number of distinct cell states to drive iXEN-like cell reprogramming.

1.5. Discussion

This study finds that Erk2 phosphorylation correlates with nuclear localisation in mouse ES cells, therefore observing Erk2 localisation dynamics with the Erk2-Venus fluorescent reporter may be

a useful means to investigate FGF/Erk mediated lineage segregation *in vivo*. Although phosphorylated Erk2 does not appear specifically nuclear localised in untreated E3.5 preimplantation mouse blastocysts, further experiments are necessary to further investigate Erk2 localisation dynamics over time, especially following FGF signalling modulation. Curiously, although serum and LIF are required to maintain mouse ES cell pluripotency, they also stimulate Erk phosphorylation in mouse ES cells (Smith et al. 2004). Previous studies demonstrate that ERK phosphorylation and nuclear translocation can be sustained or transient depending on the stimulus, and downstream outcomes can differ accordingly (Traverse et al. 1992; Adachi et al. 2002; Costa et al. 2006; Gautier et al. 2011). Thus, although both LIF and FGF stimulate Erk phosphorylation, perhaps signalling kinetics or localisation dynamics determine why only FGF-driven phosphorylation results in differentiation.

This study demonstrates that Gata6 can drive a XEN specification program independently of FGF *in vitro*, and the mechanisms of *Gata6*-mediated reprogramming may provide insights into the genetic hierarchy involved in ExEn development *in vivo*. In mouse embryos, Gata6 expression is initiated in the absence of Oct4/Nanog-mediated Fgf4-signalling, but downstream Gata4 and Sox17 expression is compromised (Frankenberg et al. 2011; Frum et al. 2013; Kang et al. 2013; Le Bin et al. 2014; Schrode et al. 2014). Similarly, FGF signalling is insufficient to drive Gata4 and Sox17 expression in *Gata6*$^{-/-}$ embryos (Schrode et al. 2014). Gata6 binding analysis reveals that Gata6 positively regulates itself as well as Fgfr2, Gata4 and Sox17. This suggests that although Gata6 is initiated independently in the embryo, it may then require a feedback loop via Fgf4/Fgfr2 and Erk signalling to reinforce its own expression to a certain threshold, which subsequently triggers the expression of Gata4 and Sox17.

This could also explain the initial co-localization of Nanog and Gata6 *in vivo,* whereby a threshold of Gata6 expression needs to be reached in order to overcome pluripotency. Gata6 overexpression *in vitro* likely exceeds this threshold, thereby leading to downregulation of

pluripotency and upregulation of ExEn genes, thus allowing iXEN reprogramming to proceed in the absence of Fgf4. Inducing high levels of Gata6 expression in the primitive endoderm of Fgf4 mutant mouse embryos would be one approach to test this hypothesis. Alternatively, Gata6 levels could be fine-tuned *in vitro* to determine if Gata6-low *Fgf4*$^{-/-}$ ES cells fail to induce Gata4 and Sox17. The doxycycline-inducible transgene system lacks the necessary resolution to address this, as titrating the dose of doxycycline has been shown to reduce the penetrance and therefore the percentage of cells inducing expression, rather than the quantitative levels within a given cell (Beard et al. 2006).

The latent levels of phosphorylated Erk observed across the mouse embryo could also reflect a regulatory threshold, as only in cells where Gata6 is independently upregulated will FGF/Erk contribute to PrE specification. Intriguingly, ERK phosphorylation sites have been identified on the GATA6 protein (Suzuki et al. 1996), and ERK-mediated activation of GATA6 has been observed in human colorectal adenocarcinoma cell lines (Adachi et al. 2008), though this is yet to be investigated in the mouse embryo. However, as Gata6 can drive iXEN reprogramming in the absence of Erk *in vitro*, this is unlikely to be the only mechanism of Gata6 activation. Nevertheless, although they may not exhibit an absolute linear relationship, FGF/Erk signalling and Gata6 expression likely eventually converge to bring about PrE specification.

Analysing global Gata6 binding during iXEN reprogramming revealed Gata6 enrichment at both pluripotency and endoderm gene targets. Gata6 therefore seemingly functions simultaneously as both an activator and repressor of genes during reprogramming, but not necessarily in eXEN cells. It has been suggested that other GATA transcription factors, such as Gata1, also have a dual activator and repressor role (Yu et al. 2009). However, it is unclear precisely how Gata6 functions to regulate both sets of target genes and whether co-factors, chromatin or other epigenetic mechanisms may influence this decision. Characterizing sequences surrounding Gata6 bound loci may reveal motifs that consistently distinguish Gata6-bound downregulated genes from

Gata6-bound upregulated genes. Furthermore, the RNA-guided CRISPR-Cas nuclease system (Cong et al. 2013; Mali et al. 2013) could be used to mutagenize Gata6 binding sites, to functionally interrogate these putative regulatory loci.

What distinguishes Gata6 from the other endoderm transcription factors tested is the speed with which it acts to induce a cell fate switch in the absence of selection. Indeed, most reprogramming, including iPS cells, can take several days of selective culture. As the induction of Sox17 takes over 2 weeks to downregulate pluripotency (McDonald et al. 2014), this suggests that indirect mechanisms eventually lead to XEN conversion, perhaps via Gata6. However, Sox17 has been shown to bind to some extraembryonic endoderm genes (Niakan et al. 2010), and thus likely directly regulates their expression. It would therefore be interesting to determine if Sox17 can induce XEN reprogramming in the absence of Gata6, though seems unlikely given that *Gata6*$^{-/-}$ mouse ES cells fail to initiate cXEN cell conversion in growth factor mediated conditions, while *Sox17*$^{-/-}$ cells generate, but fail to maintain, cXEN cells (Cho et al. 2012).

The Gata6-overexpression system also represents a useful tool to further analyse the functional requirement for genes within the ExEn network. A recent study demonstrated that Sox7 overexpression was insufficient to drive iXEN reprogramming (Kinoshita et al. 2015), consistent with the results presented here. Interestingly, Gata6 overexpression in *Sox7*$^{-/-}$ mouse ES cells was sufficient to drive a morphological conversion to XEN-like cells, and lack of Sox7 did not affect upregulation of endoderm genes. Moreover, *Sox7*$^{-/-}$ cells contributed to ExEn-derivatives in embryoid bodies (Kinoshita et al. 2015). This would suggest that despite its expression being restricted to the PrE (Artus et al. 2011), Sox7 seems dispensable for the establishment of this lineage. However, it has been hypothesized that Sox7 and Sox17 have overlapping functions in the PrE (Kanai-Azuma et al. 2002; Artus et al. 2011), so Sox17 expression may compensate for the lack of Sox7 in this instance.

Remarkably, Gata6 drives iXEN cells from mouse neural stem cells, showing that it is a broad inducer of reprogramming and can overcome intrinsic programs within distinct cell types. However, it would be surprising if all endoderm target genes remained readily accessible for Gata6 direct regulation once cells had acquired a neural fate. Therefore, it is possible that Gata6 may be functioning as a pioneering transcription factor, exposing otherwise closed heterochromatic regions as has been shown for FoxA2 and Gata4 (Zaret and Carroll 2011). In addition, Gata6 has also been shown to be required for endoderm and mesoderm derived cell types (Xin et al. 2006; Tiyaboonchai et al. 2017), so it would be curious to determine whether Gata6 overexpression consistently drives one reprogramming outcome. Further characterization of Gata6-mediated reprogramming in several cellular contexts would allow interrogation of the relationship between transcription factors, signalling and epigenetics in driving cell state transitions.

Finally, this study reveals that GATA6 induction can initiate ExEn expression in human ES cells, and is also sufficient to inhibit core pluripotency gene expression, distinguishing it from previous studies with alternative ExEn transcription factors *SOX7* or *SOX17* (Seguin et al. 2008). It would be interesting to determine if a longer GATA6 induction may be effective in capturing stable human iXEN cells, or if alternative conditions may be required. In addition, recent transcriptomic analysis of human embryos (Blakeley et al. 2015) may lead to the identification of signalling pathways that may be important to stabilize GATA6-induced human XEN cell lines. Comparing gene expression in *GATA6*-transduced cells to expression data from the human PrE (Blakeley et al. 2015) and definitive endoderm (Loh et al. 2014) would also be a useful indicator as to whether GATA6 is truly driving an ExEn identity. Altogether this demonstrates that Gata6 is a versatile and potent reprogramming factor that can act alone to drive a cell fate switch from diverse cell types.

2. FGF signalling and lineage specification in early human development

In contrast to its role in promoting differentiation in the mouse Epi and in mouse ES cells, FGF is routinely added to maintain human ES cell pluripotency (Thomson et al. 1998; Amit et al. 2000; Reubinoff et al. 2000; Xu et al. 2001; Cowan et al. 2004; Levenstein et al. 2006; Ludwig et al. 2006; Chen et al. 2011). Human ES cells also require Activin/Nodal signalling to maintain pluripotency and unlike the mouse, are not dependent on LIF or STAT3 (Thomson et al. 1998; Daheron et al. 2004; Humphrey et al. 2004; Vallier et al. 2005; Vallier et al. 2009). FGF/MAPK inhibition in human ES cells affects Nanog expression and promotes neural differentiation (Greber et al. 2010; Greber et al. 2011), which is consistent with a role for FGF in maintaining pluripotency.

Curiously, inhibiting MAPK signalling in human blastocysts has no effect on pluripotency (or ExEn) gene expression (Kuijk et al. 2012; Roode et al. 2012). Previous studies indicate that FGF receptors are not appreciably expressed in the human blastocyst although they are present in human ES cells (Dvorak et al. 2005; Kunath et al. 2014); perhaps this differential expression explains the contrasting sensitivity to FGF signal modulation. Alternatively, FGF may function via an Erk-independent pathway in the human embryo.

Only a few studies thus far have interrogated signalling in the human embryo, likely partly due to the relative scarcity of embryos surplus to family building via IVF that are available for research. The initial investigations that established the conditions for human ES cell derivation were carried out in non-human primates (Thomson et al. 1995; Thomson et al. 1996), and may not necessarily reflect the requirements of early human embryonic development. Consequently, modulating signalling and analysing gene expression patterns in the human Epi directly would help clarify the role of FGF in the embryo and shed light on these incongruities.

Therefore, in this chapter, I investigate the effect of stimulating FGF signalling during early human embryonic development, and also explore whether additional alternative signalling pathways are required to maintain the human Epi.

2.1. Exploring FGF signalling in the human embryo

The incongruous effects of FGF signalling modulation in the human blastocyst and ES cells imply that there may be distinct pathway components acting in these contexts. Our lab has recently compiled a single cell transcriptomic dataset from human embryos at multiple developmental stages (1-cell zygote to blastocyst), collating newly collected samples and data from a recently published study (Yan et al. 2013; Blakeley et al. 2015). Therefore, this dataset represented a useful resource for exploring human blastocyst-specific signalling pathways. In addition, data from single cell analysis of human ES cells was also collated for comparison with the Epi compartment.

Paul Blakeley performed analysis of the sequencing data. Briefly, the reads per kilobase of exon model per million mapped reads (RPKM) method (Mortazavi et al. 2008) was used to normalize for sequencing depth and transcript length. The threshold of significance was set at >5 RPKM for median gene expression, which has been shown to reliably capture functional mRNA expression (Hebenstreit et al. 2011). To investigate differences in global gene expression, a PCA of the top 8000 genes with the most variable expression was performed on the human late blastocyst samples, which clustered into 3 groups (Blakeley et al. 2015). Cells were then segregated into Epi, PrE or TE depending on enrichment of lineage-associated genes in each cluster, as determined by two independent analyses, NOISeq and DESeq (Anders and Huber 2010; Tarazona et al. 2011). Considerable overlap was observed in the genes predicted by these two independent statistical methods (Blakeley et al. 2015), lending confidence to the resulting lineage assignations. Finally, boxplots of RPKM values were generated to show the range of gene expression in human ES cells and the blastocyst samples.

2.1.1. FGF signalling-related transcripts are differentially expressed in existing human ES cells and human blastocysts

Interrogating the RPKM boxplot dataset identified a number of differences in FGF receptor and ligand transcript expression between the various cell types. *FGFR1* was expressed in both the Epi and human ES cells, but most enriched in the PrE, while *FGFR2* was expressed only just above the >5 RPKM significance threshold in human ES cells, the Epi and PrE, and not significantly in the TE (Figure 2.1A). In contrast, *FGFR3* and *FGFR4* median expression fell below the >5 RPKM significance threshold in both human ES cells and the Epi, and the receptors were respectively more abundant in the PrE, or both PrE and TE (Figure 2.1A).

Previous studies suggesting that FGF receptors are not appreciably expressed in the human blastocyst had analysed receptor expression in whole blastocysts via RT-PCR (Kunath et al. 2014). The boxplot dataset indicates *FGFR1* is most enriched in the PrE, which makes up a relatively small proportion of the embryo in the late blastocysts analysed here, therefore perhaps single cell analysis allows for greater resolution by compensating for an overabundance of non-expressing cells in whole-blastocyst samples. Furthermore, FGFR1 protein has been detected in the human embryo (Niakan and Eggan 2013), with an initially broad expression pattern that becomes restricted to a few cells at later stages. *FGF2* ligand transcripts were not appreciably expressed in the epiblast, while median *FGF4* expression was below 5 RPKM (Figure 2.1B). In contrast, only *FGF2* was appreciably expressed in human ES cells. A number of additional FGF ligands were also detected, mainly in human ES cells (Figure 2.1B).

Analysing downstream kinase expression in the single cell dataset revealed expression of MAPK signalling components across all the blastocyst lineages, including *MAP2K1 (MEK1)* and *MAPK1 (ERK2)* (Figure 2.2A). Interestingly, the Ras gene *ERAS*, a novel MAPK signalling component thought to promote growth in mouse ES cells (Takahashi et al. 2003) was not appreciably expressed in human ES cells or the epiblast. There appeared to be no preferential

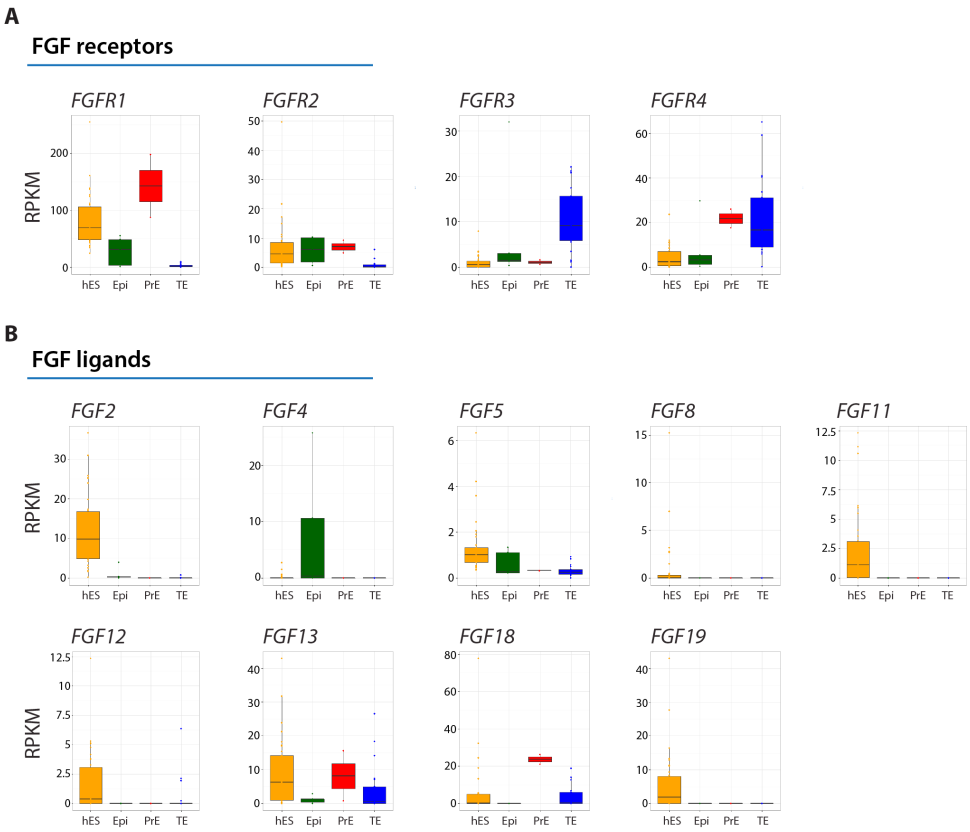


Figure 2.1. FGF signalling pathway components in the early human blastocyst and ES cells. Box plots of RPKM data showing the range of gene expression in human ES cells (yellow), or Epi (green), PrE (red) or TE (blue) blastocyst lineages based on gene expression patterns. Boxes correspond to the first and third quartiles, horizontal line to the median, whiskers extend to 1.5 times the interquartile range, and dots denote outliers. Plots are shown for **(A)** FGF receptors, and **(B)** FGF ligands whose expression was detected in at least one blastocyst lineage or in ES cells. Box plots generated by Paul Blakeley. n = 23 hES, 5 Epi, 2 PE and 23 TE cells.

enrichment compared to other lineages that would suggest a requirement for FGF signalling in the Epi or human ES cells. If FGF does indeed function via a distinct pathway in the embryo, one potential alternative is the phosphatidylinositol 3-kinase (PI3K) pathway, which has been implicated in mouse ES cell self-renewal (Burdon et al. 1999; Paling et al. 2004). A number of PI3K signalling kinases were also detected in the human blastocyst (Figure 2.2B). In all, there appeared to be evidence for a capability for FGF signal transduction within the embryo, at least based on transcript availability. Therefore, experiments were set up to treat human embryos with exogenous FGF and determine the effect on lineage segregation.

2.1.2. FGF stimulation does not enhance the pluripotent epiblast

Previous experiments in mouse embryos had identified a dose-dependent effect of FGF stimulation on cells within the ICM (Yamanaka et al. 2010). FGF treatment resulted in a cell fate conversion of Nanog-expressing Epi cells to Gata6-expressing PrE cells, with doses from 250 ng to 1 μg/ml FGF supplemented with 1 μg/ml heparin converting all cells in the E4.5 mouse ICM to PrE (Yamanaka et al. 2010). Therefore, a similar treatment schedule was set up with human embryos to investigate the effect of FGF stimulation (Figure 2.3A); if FGF was required for pluripotency expression, the converse effect should be observed.

8-cell human embryos (E2.5) were first cultured in media supplemented with 1 μg/ml FGF2 and 1 μg/ml heparin (Figure 2.3B), similar to the mouse embryo experiments. By E6.5, untreated control embryos formed an expanded blastocyst and expressed both NANOG and GATA6 (Figure 2.3B). In contrast, FGF-treated embryos did not express NANOG and appeared smaller compared to sibling-matched controls, suggesting they may be developmentally delayed (Figure 2.3B). Additionally, cellular debris was observed within some embryos (arrows), perhaps indicative of apoptotic cells (Figure 2.3B). This differs from FGF-treated mouse embryos, which appear developmentally normal despite lineage conversion within the ICM (Yamanaka et al. 2010).

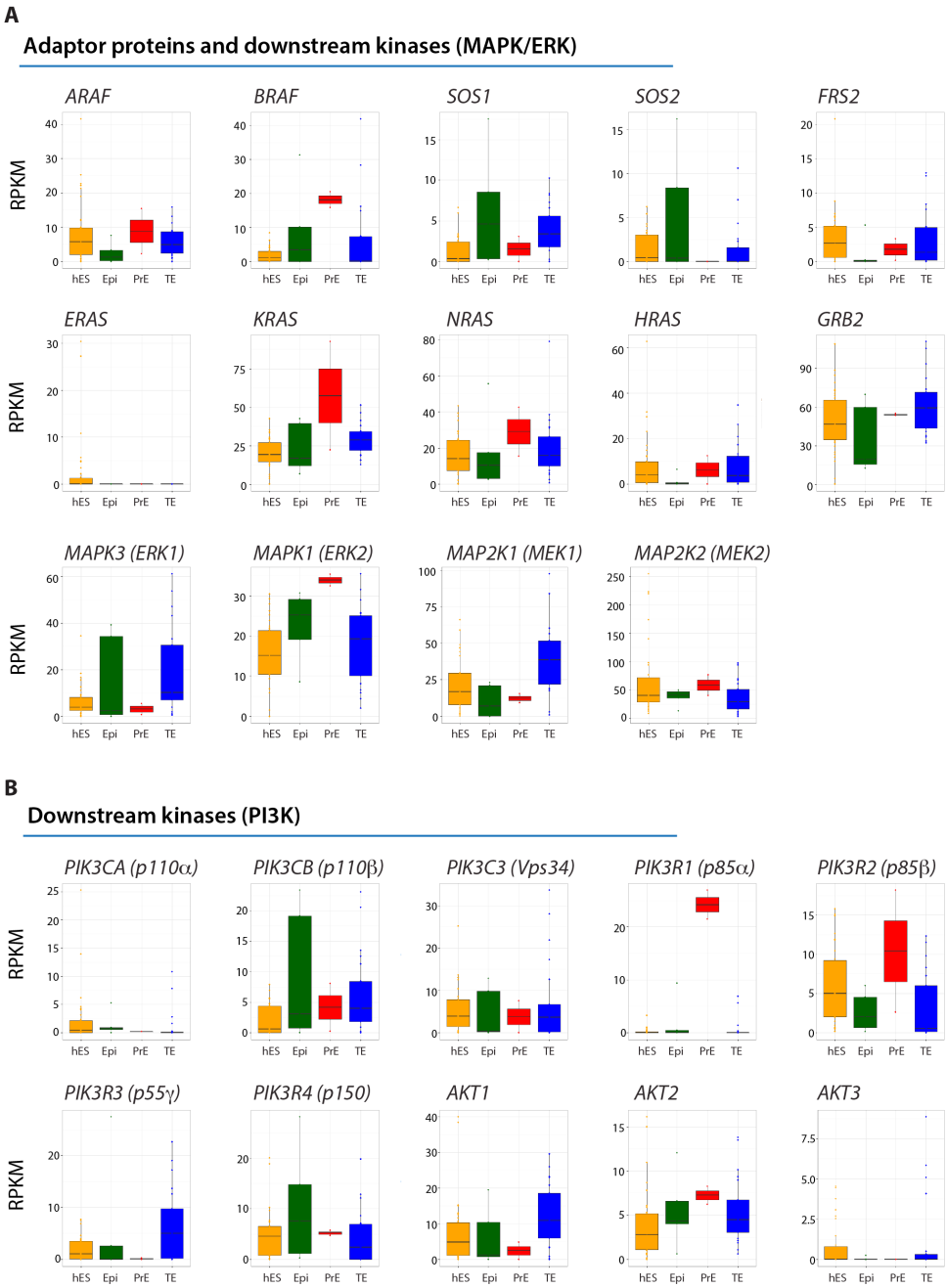


Figure 2.2. Downstream components of FGF signalling are present in the human blastocyst. Box plots of RPKM data showing the range of gene expression in human ES cells (yellow), or Epi (green), PrE (red) or TE (blue) blastocyst lineages based on gene expression patterns. Boxes correspond to the first and third quartiles, horizontal line to the median, whiskers extend to 1.5 times the interquartile range, and dots denote outliers. Plots are shown for **(A)** MAPK/ERK and (**B**) PI3K signalling components. Box plots generated by Paul Blakeley. n = 23 hES, 5 Epi, 2 PE and 23 TE cells.

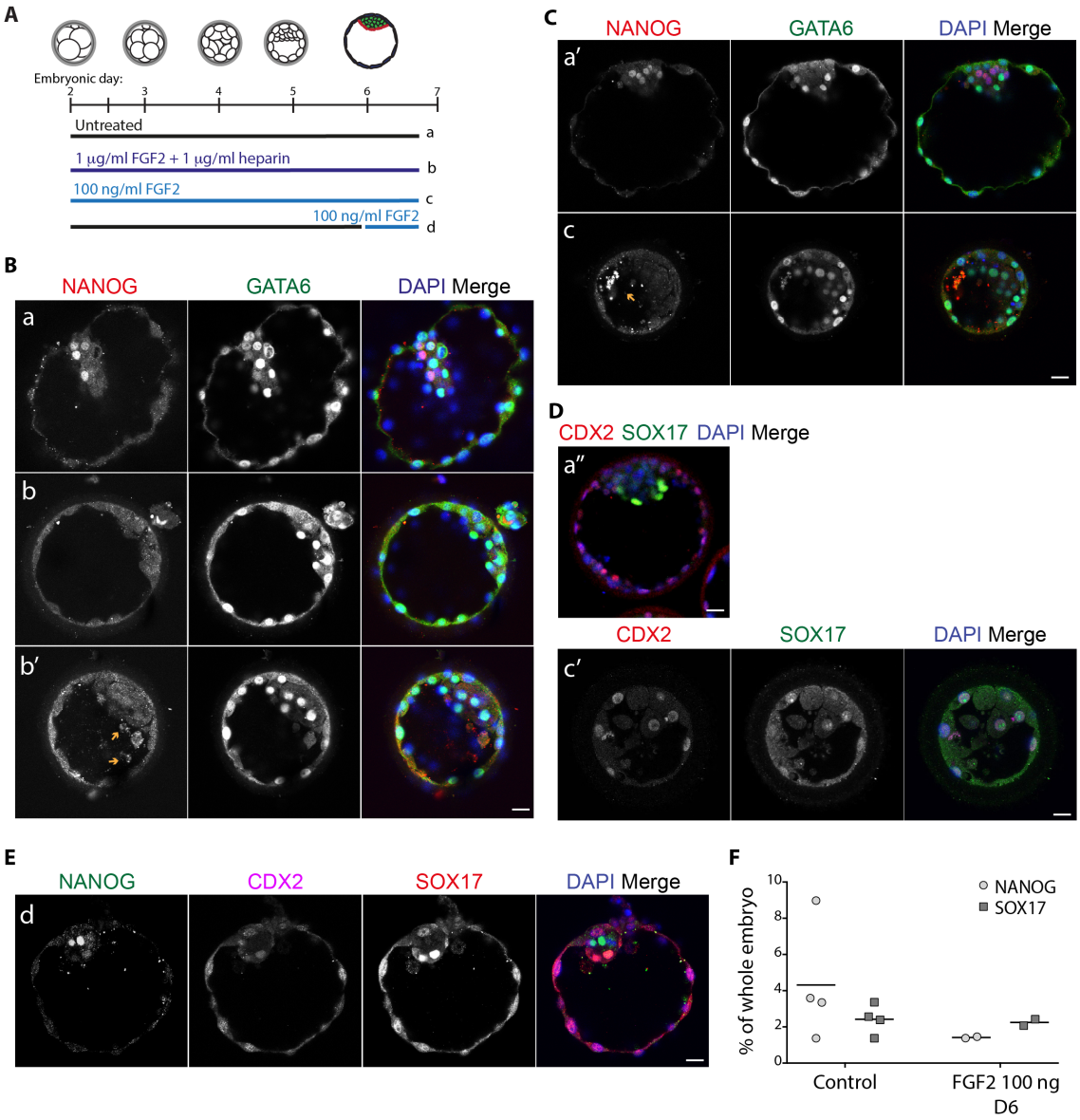


Figure 2.3. FGF treatment negatively affects establishment of the pluripotent Epi. (A) Treatment schedule. **(B)** Representative images of immunofluorence analysis for NANOG (red) and GATA6 (green) with DAPI merge (blue) at E6-7 in an untreated control embryo (a) or in embryos treated with 1 µg/ml FGF2 and 1 µg/ml heparin from E2.5 (b, b′). Orange arrows indicate cell debris. 4 control and 3 FGF2/heparin-treated embryos were analysed. **(C, D)** Representative immunofluorescence analysis for NANOG (red) and GATA6 (green), or CDX2 (red) and SOX17 (green), with DAPI merge (blue) at E6.5 in control embryos (a′, a″) or embryos treated with 100 ng/ml FGF2 from E2.5 (c, c′). Orange arrows indicate cell debris. 2 control and 4 FGF2-treated embryos were analysed. **(E)** Representative immunofluorescence analysis for NANOG (green), CDX2 (magenta) and SOX17 (red), with DAPI merge (blue) at E7 in an embryo treated with 100 ng/ml FGF2 from E6 (c, c′). 2 embryos were analysed. **(F)** Proportions of NANOG and SOX17-expressing cells in control or FGF2-treated embryos as a percentage of total cell number as determined by MINS. Images taken by Kathy Niakan. 4 control and 2 FGF2-treated embryos were analysed.

Embryos were next treated with 100 ng/ml FGF2, a concentration commonly used to maintain human ES cell pluripotency *in vitro* (Ludwig et al. 2006), and thus possibly more relevant in a human context. Again, by E6.5 treated embryos lacked NANOG-expressing cells but retained GATA6-expressing cells, with indications of potentially apoptosed cells (Figure 2.3C). Furthermore, although these embryos still expressed the TE marker CDX2, there appeared to be unusually extensive co-localisation of CDX2 and the PrE marker SOX17 (Figure 2.3D), which suggests there may also be an effect on these lineages despite the continued expression of key lineage markers. However, FGF might instead be required for maintenance of the epiblast once established, promoting proliferation or survival for example, and perhaps the observed adverse effect resulted from FGF exposure at a premature developmental stage.

To investigate this, embryos were treated with 100 ng/ml FGF2 from E6, by which time NANOG and OCT expression is restricted to the ICM (Niakan and Eggan 2013). Treated embryos appeared developmentally normal and NANOG expression was detected in the ICM (Figure 2.3E). To determine whether the proportion of NANOG-expressing cells was affected, the automated software tool MINS 1.3 (Lou et al. 2014) was used to detect and segment DAPI-stained nuclei. The number of NANOG- or SOX-17 expressing cells was then calculated as a proportion of the total cells in the embryo, and treated embryos then compared to controls to determine any statistically significant changes in proportion. Although the proportion of NANOG- or SOX17-expressing cells in FGF-treated embryos initially appeared lower than expected, values were within the range of controls (Figure 2.3F), and further embryos are being analysed to confirm this. Altogether, this suggests that in contrast to its pluripotency maintenance function in human ES cells, FGF is detrimental to the establishment of the pluripotent epiblast compartment *in vivo*. Treatment at later stages, while no longer detrimental, does not seem to have a detectable effect on lineage proportions within the blastocyst. This implies that rather than operating via an alternative pathway in the human embryo, FGF is unlikely to be required at all.

2.2. Characterising gene expression in the human Epi

It is increasingly apparent that existing human ES cells are generally distinct from the human Epi in terms of their gene expression patterns (Blakeley et al. 2015), and the discrepancies observed with FGF treatment seemingly support this. Therefore, analysing differential gene expression between lineages in the human blastocyst could reveal key pathways that could then be manipulated to investigate how the Epi is established and maintained. Furthermore, modulating these pathways to adapt existing human ES cells and for new derivations could define culture conditions to support human ES cells that were more similar to, and thus a better model for, the originating human Epi compartment.

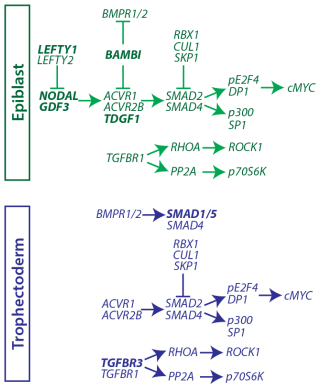
To interrogate lineage-specific gene expression patterns in the human blastocyst, the DESeq differential gene expression dataset (Blakeley et al. 2015) was used. Comparing genes enriched in the human Epi vs. TE, and Epi vs. PrE, identified classic factors such as *NANOG* and *SOX2* enriched in the Epi, *CDX2* and *GATA3* in the TE, and *GATA4* and *SOX17* in the PrE (Figure 2.4A). Several additional genes were identified as enriched in each lineage, such as *ESRG* (*HESRG*), which has been signposted as a putative pluripotency gene (Zhao et al. 2007; Li et al. 2013). The Kruppel-like factor family member *KLF17* (van Vliet et al. 2006) was also identified as enriched in the Epi (Figure 2.4A), and boxplot RPKM data revealed it was exclusive to this lineage (Blakeley et al. 2015). If these novel factors also segregated with known lineage-associated genes at the protein level, it would validate the ability of the transcriptomic analysis to predict novel lineage-associated factors. Immunofluorescence staining confirmed that KLF17 protein expression was also restricted to the human Epi, and was coincident with NANOG (Blakeley et al. 2015). Furthermore, the human PrE-specific transcript *FOXA2* was also validated at the protein level (Blakeley et al. 2015), although it is only associated with later stage endoderm in the mouse (Kimura-Yoshida et al. 2007). Consequently, the differential expression transcriptome dataset appeared to be a valid predictive tool.

A

	Gene ID	Epi mean	TE mean	Log2 FC
Epi enriched	*LEFTY1*	914.6	0.1	-12.6
	EPHA4	2242.7	0.7	-11.6
	NODAL	1446.1	1.2	-10.3
	NANOG	6741.3	9.7	-9.4
	KLF17	2043.7	9.2	-7.8
	IL6R	4287.3	20.1	-7.7
	GDF3	6339.4	38.3	-7.4
	SOX2	1100.3	11.1	-6.6
TE enriched	*PTGES*	0.4	461.2	10.3
	PLAC8	0.5	142.4	8.2
	GATA2	16.2	4528.6	8.1
	CDX2	2.5	431.9	7.4
	TEAD3	3.4	281.6	6.3
	TGFBR3	100.5	3893.7	5.3
	GATA3	255.1	8457.3	5.1
	SMAD5	528.8	2252.0	2.1

	Gene ID	Epi mean	PrE mean	Log2 FC
Epi enriched	*PRDM14*	1483.5	0.5	-11.6
	NANOGP1	2336.2	1.5	-10.6
	NODAL	1174.6	1.0	-10.2
	GDF3	5103.2	4.4	-10.2
	NANOG	5534.3	29.5	-7.5
	ESRG	4234.7	34.1	-6.9
	KLF17	1664.8	14.4	-6.8
	TDGF1	9084.4	133.7	-6.1
PrE enriched	*IGF1*	1.2	3792.4	11.6
	ROBO2	0.2	517.6	11.2
	FOXA2	1.1	1458.4	10.4
	COL4A2	2.8	2838.5	10.0
	BMP6	2.3	2370.3	10.0
	GATA4	5.3	3515.5	9.4
	SOX17	1.0	537.9	9.0
	PDGFRA	123.3	3139.0	4.7

B



C

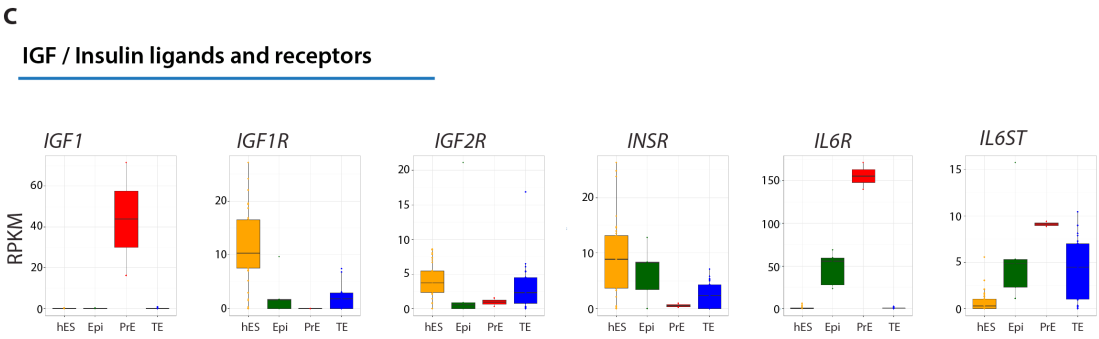


Figure 2.4. Characterising gene expression in the human Epi. (A) DESeq analysis indicating selected genes significantly differentially expressed in the Epi compared to the TE or PrE. The mean gene expression and log2 fold change (FC) are shown. **(B)** Summary of TGFβ signalling components expressed at an RPKM value >5 in human EPI or TE. Differentially expressed genes as determined by DESeq analysis are noted in bold. **(C)** Box plots of RPKM data showing the range of gene expression for insulin/IGF1 signalling pathway receptors and ligands. Box plots generated by Paul Blakeley. n = 23 hES, 5 Epi, 2 PE and 23 TE cells.

DESeq analysis identified a number of TGFβ/Nodal signalling-associated genes as enriched in the Epi, including the ligands *NODAL* and *GDF3*, the EGF-CFC co-receptor *TDGF1* (prior annotation *CRIPTO*), and the antagonists *LEFTY1* and *BAMBI* (Fig. 2.4A, 2.4B). Additional TGFβ signalling pathway components were present in the Epi above the RPKM >5 significance threshold, but were not differentially expressed as they were also expressed in the TE (Fig. 2.4B; (Blakeley et al. 2015). However, distinct upstream components *TGFRB1* and *SMAD5* were enriched in the TE (Fig. 2.4A, 2.4B), which are instead involved in negative regulation of the pathway, suggesting that TGFβ/Nodal is differentially activated in the Epi compared to the TE. In addition, the insulin/IGF1 signalling pathway ligand *IGF1* was identified as enriched in the PrE, while the IL6R receptor subunit *IL6R*, which can also bind insulin ligands , was enriched in the Epi (Fig. 2.4A).

To identify putative signalling pathways on a wider scale, the GOrilla gene ontology (GO) analysis tool (Eden et al. 2007; Eden et al. 2009) was used to mine the dataset of genes expressed >5 RPKM in half or more of the Epi cells analysed. This would allow for inclusion of downstream pathway components that may be present in the Epi but not necessarily be differentially expressed. A number of signalling related categories were identified, with several linked to regulation of Wnt and TGFβ signalling (Appendix Table X1). Analysing the equivalent TE dataset also pulled out TGFβ signalling related terms, consistent with the DESeq data above (Appendix Table X2), as well as several Wnt related terms. This may indicate why modulating Wnt signalling in existing human ES cells has alternately been found to promote either pluripotency or differentiation (Sato et al. 2004; Davidson et al. 2012).

Both TGFβ/Nodal and insulin/IGF1 pathways have been previously implicated in either human or mouse ES cell culture systems. TGFβ/Nodal signalling has been shown to regulate pluripotency gene expression and is routinely used to maintain human ES cells (Amit et al. 2004; Vallier et al. 2004; Vallier et al. 2005; Brown et al. 2011). Addition of the TGFβ family ligand

Activin A is also sufficient to maintain human ES cell pluripotency and Activin is expressed and secreted by the MEF layer used in classical human ES cell culture (Beattie et al. 2005; James et al. 2005; Vallier et al. 2005). Nodal-overexpressing cells can also be maintained in chemically defined medium (CDM) without FGF or serum (Vallier et al. 2004). Culture in Activin alone supported more OCT4-positive colonies than Nodal alone, though FGF supplementation in tandem with NODAL rescued this deficit. However, higher doses of Nodal can dispense with FGF entirely (Vallier et al. 2005).

Insulin/IGF1 signalling has been implicated in self-renewal in alternative contexts (Malaguarnera and Belfiore 2014), and may also have a role in maintaining human ES cells (Bendall et al. 2007; Wang et al. 2007). Revisiting the boxplot RPKM dataset revealed that neither *IGF2* nor insulin (*INS*) ligands are detectable in either the blastocyst or human ES cells (Fig. 2.4C), but *IGF1* is indeed enriched in the PrE (Fig. 2.4C), as indicated in the DESeq analysis. *IGF1R* was expressed in human ES cells but fell below the 5 RPKM significance threshold in the blastocyst lineages, while the insulin receptor (*INSR*) was detected in the Epi and in human ES cells (Fig. 2.4C). *IGF2R* was detected at below threshold levels in both human ES cells and the blastocyst lineages.

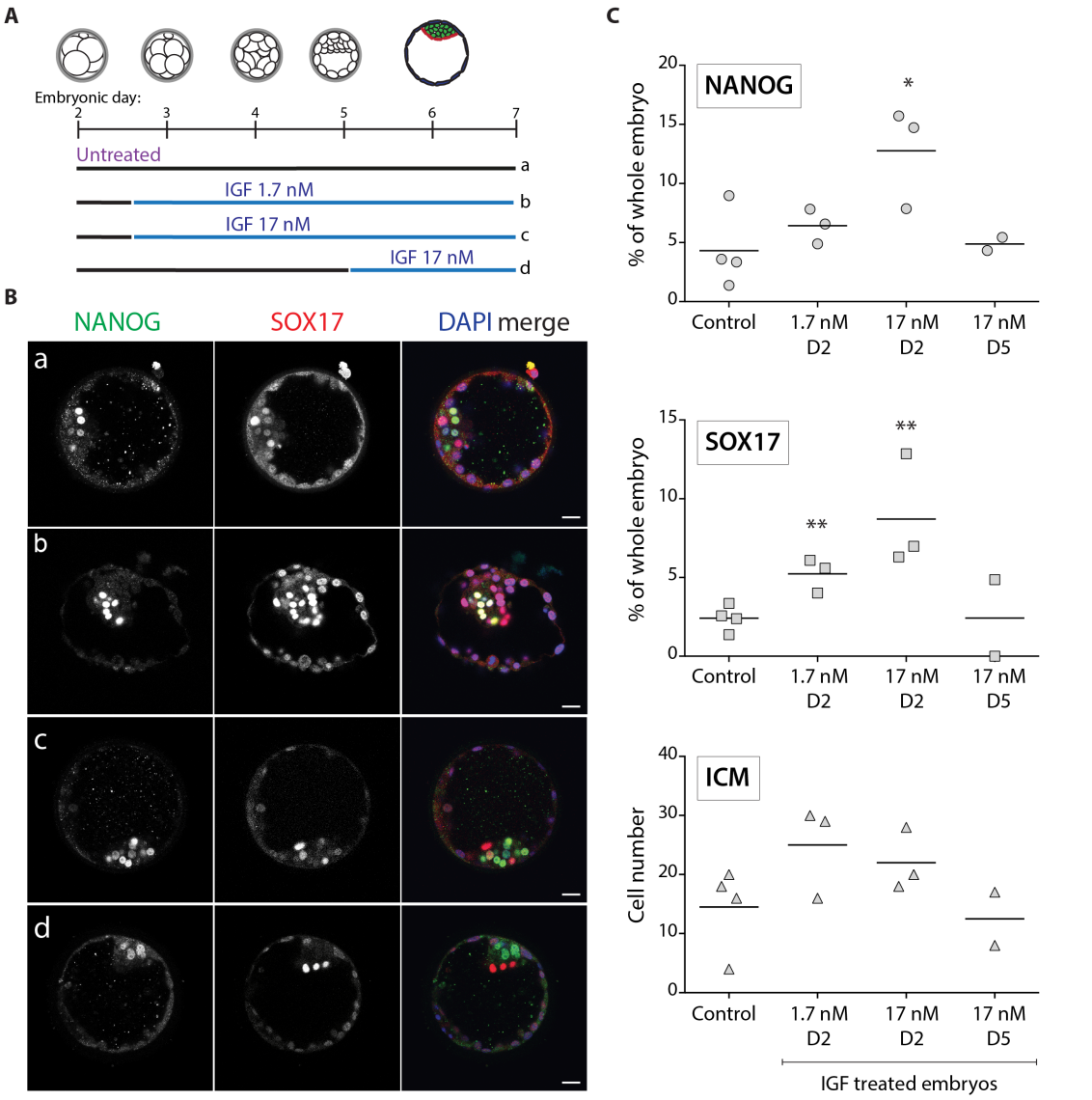
In addition, the interleukin-6 receptor *IL6R*, shown to crosstalk with insulin/IGF1 signalling in an alternative context (Abroun et al. 2004), was highly enriched in both the Epi and PrE, while its co-receptor *IL6ST* (formerly *GP130*) was present in the PrE, and just below the threshold in the Epi. Altogether, this suggests a possible role for the insulin/IGF1-signalling pathway in the Epi and PrE lineages, though it is unclear what effect if any this may have in terms of lineage determination. Therefore, the TGFβ/Nodal and insulin/IGF1 signalling pathways seemed ideal starting points for signalling modulation in the human blastocyst.

2.2.1. IGF treatment promotes proliferation within the human ICM

In addition to its proposed role in human ES cell maintenance (Bendall et al. 2007), insulin/IGF1 signalling has also been implicated in embryo development. Previous studies suggest that IGF1 may be present in the Fallopian tube and uterine fluid (Lighten et al. 1997; Lighten et al. 1998), and studies treating human embryos with IGF1 identified an increase in proportion of embryos developing to the blastocyst stage, as well as an increase in the number of cells in the ICM compartment (Lighten et al. 1998; Spanos et al. 2000). Similar effects had been observed in mouse embryos in earlier experiments (Harvey and Kaye 1992). However, these studies did not distinguish between Epi and PrE cells within the ICM.

Putative treatment factor concentrations were initially tested on 4 – 8-cell mouse embryos to determine toxicity, as measured by the number of treated embryos developing to the blastocyst stage relative to controls. If concentrations were deemed non-toxic, human embryos were then cultured from either E2.5 (cleavage) or E5 (early blastocyst) till E6-7 (late blastocyst) (Fig. 2.5A). Treated and control embryos were stained for NANOG and SOX17 to mark Epi and PrE lineages respectively, and thus determine whether a given treatment affected a particular ICM lineage or induced a cell fate switch. NANOG and SOX17 were chosen as they are well-established markers of these two lineages, but in future newly identified human-specific factors such as KLF17 and FOXA2 (Blakeley et al. 2015) may be additional factors to consider. The proportions of Epi and PrE cells were quantified using the MINS software tool.

8-cell human embryos (E2.5) were first treated with 1.7 nM IGF1, the concentration found to affect ICM cell number in earlier studies (Lighten et al. 1998). By E6.5, treated embryos had formed an expanded blastocyst and expressed both NANOG and SOX17 (Fig 2.5B). MINS analysis identified a slight increase in the proportion of NANOG-expressing cells, although this was not significantly different compared to controls (Fig 2.5C, 2.5D), while the proportion of SOX17-expressing cells had significantly increased (Fig 2.5C, 2.5D). The total number of ICM



D

Condition	Nanog	Nanog %	Sox17	Sox17 %	ICM	ICM %	Total cell no.
Control	8.5 ± 4.9	4.3 ± 3.2	5.75 ± 3.9	2.4 ± 0.8	14.5 ± 7.2	7.0 ± 3.6	205.5 ± 63.4
IGF 1.7 nM D2 - D6/7	13.7 ± 4.5	6.4 ± 1.5	11.3 ± 4.5	5.2 ± 1.0 *	25.0 ± 7.8	11.7 ± 1.9	225 ± 95.5
IGF 17 nM D2 - D6/7	13.3 ± 4.9	12.8 ± 4.3*	8.7 ± 0.6	8.7 ± 3.6 **	22.0 ± 5.3	21.5 ± 7.2	108.7 ± 31.2*
IGF 17 nM D5 - D6/7	8.0 ± 0	4.9 ± 0.8	4.5 ± 6.4	2.4 ± 3.4	12.5 ± 6.4	7.3 ± 2.6	166.0 ± 46.3

Figure 2.5. IGF treatment promotes proliferation within the human ICM. **(A)** Treatment schedule **(B)** Immunofluorescence analysis for NANOG (green) and SOX17 (red), with DAPI merge (blue) at E6-7 in control embryos (a) or embryos treated with IGF1 at 1.7 nM or 17 nM from either E2.5 or E5 as indicated in (A). Scale bars: 100 µm. 4 control and 3 embryos per IGF treatment condition were analysed. **(C)** Proportions of NANOG and SOX17-expressing cells in control or treated embryos as a percentage of total cell number as determined by MINS. Total ICM cell number also plotted. 4 control and 2 or 3 embryos per IGF treatment condition were analysed. **(D)** Cell counts and calculated proportions used to generate graphs in (C). * $p < 0.05$, ** $p < 0.01$.

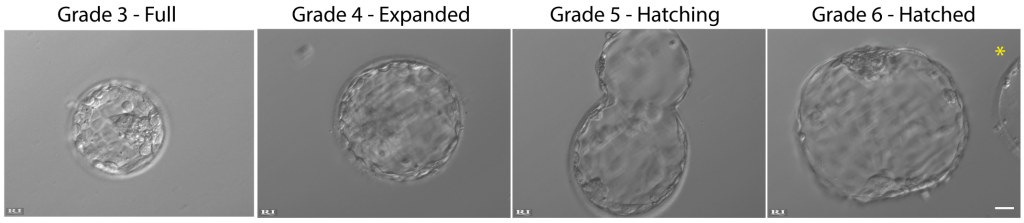
cells had indeed increased slightly relative to controls, though there was no difference in total number of cells in the embryo (Fig 2.5C, 2.5D), consistent with previous work (Lighten et al. 1998).

Embryos were next treated with a 10-fold higher IGF1 concentration to determine if a dose-response effect could be observed. At E6.5, embryos treated with 17 nM IGF1 had significantly increased proportions of both NANOG- and SOX17-expressing cells (Fig 2.5B, 2.5C, 2.5D), and total ICM cell number was again slightly increased (Fig 2.6C, 2.6D). Embryos were then treated with 17 nM IGF1 from E5, to determine if this effect was conserved with IGF1 exposure at the later stage. However, in these later-stage embryos proportions of both NANOG- and SOX17-expressing cells, and total ICM cell number, remained equivalent to controls (Fig 2.6C, 2.6D), suggesting a longer exposure coupled with higher dosage is required for this effect.

While calculating the proportion data, it was noted that there were significantly fewer total cells in the embryos treated with 17 nM from E2.5 (Fig 2.5D). Although this may indicate that the expansion of the ICM is occurring at the expense of the TE, it may also reflect an unavoidable element of human embryo culture. Even though embryos were selected from the same developmental stage and treated for the same temporal period after thawing, human embryos are frozen at differing times and thus develop according to slightly different timetables. To determine whether this may explain the lower total cell number, the Gardner and Schoolcraft embryo grading scale (Gardner and Schoolcraft 1999) was used to assess both treated and control embryos.

Grading revealed that when collected for analysis, more embryos in the control and 1.7 nM datasets were fully hatched (grade 6), or hatching (grade 5), from the zona compared to embryos treated with 17 nM, which were mainly full (grade 3) or expanded (grade 4) (Fig 2.6A). Given the small sample size, this difference would likely have a greater impact on analysis. Increasing the

A



Condition	Grade 3	Grade 4	Grade 5	Grade 6
Control	-	2	2	-
IGF 1.7 nM D2 - D6/7	-	1	1	1
IGF 17 nM D2 - D6/7	1	-	1	-
IGF 17 nM D5 - D6/7	1	2	-	-

B

Condition	Thaw stage	Cultured	Blastocysts recovered	% recovery
Control*	D2	108	29	26.9
IGF 1.7 nM	D2	9	5	58.3
IGF 17 nM	D2	15	3	20.8

*Data from annual summary (2015 - 2016) of total human embryo thaws in Niakan lab

Figure 2.6. Blastocyst grading and recovery rates. (A) Blasotcyst expansion grades for control and IGF-treated embryos. Representative images are included for each classification. Scale bars: 100 μM. Yellow asterisk in final panel indicates remnant zona pellucida just out of view. **(B)** Blastocyst recovery rates for IGF-treated embryos in this study compared to control data from lab annual records. Blastocysts were recovered at E6 - E7.

number of both control and treated embryos would determine if there are consistently more expanded embryos in control embryos, which may indicate a TE-specific effect of IGF treatment. In addition, a larger pool would hopefully encompass embryos at various expansion stages, ideally with a great enough number to allow for binning according to grade and subsequent comparisons without loss of statistical power. Furthermore, analysing TE protein expression would elucidate whether this lineage is being appropriately specified.

Finally, examining the rates of blastocyst development between treated and control embryos indicated that IGF1 exposure at 1.7 nM doubled the number of E2.5 blastocysts that successfully formed blastocysts (Fig 2.6B). This effect was not observed at the higher concentration of IGF1 (Fig 2.6B). This may be especially relative in an IVF context, where lower rates of embryo progression to blastocyst compared to mouse embryo culture may reflect a sub-optimal, though reasonably adequate, human embryo culture environment. Furthermore, it has been hypothesised that larger ICMs positively correlate with implantation rates following blastocyst transfer (Richter et al. 2001; Lagalla et al. 2015). In all, IGF1 treatment has a demonstrable proliferative effect in the human embryo, specifically in the ICM, and may also positively promote embryo survival *in vitro*.

2.2.2. Nodal signalling is required to maintain the pluripotent epiblast

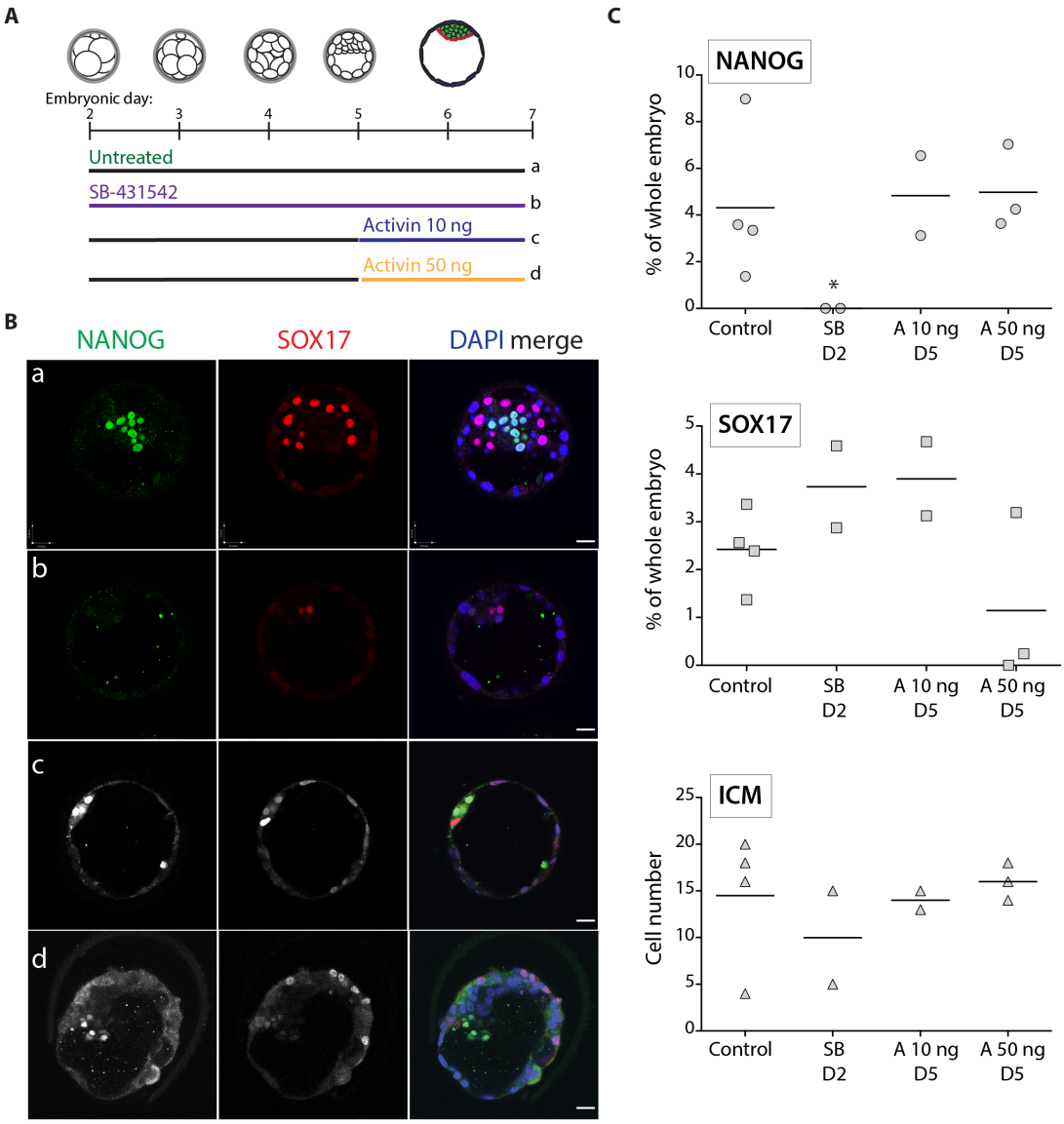
To investigate whether TGFβ/Nodal signalling is active during early human development, embryos were first treated (Norah Fogarty) with SB-431542 (SB), a potent inhibitor of TGFβ/Activin-induced Smad2 phosphorylation that had been shown to reduce NANOG expression in human ES cells (Callahan et al. 2002; Inman et al. 2002; Greber et al. 2010; Blakeley et al. 2015). Embryos were cultured with 40 μM SB, a concentration shown to be non-toxic and effective at blocking Nodal signalling in peri-implantation mouse embryos (Granier et al. 2011). In comparison, SB-treated pre-implantation mouse embryos retained Nanog expression

(Blakeley et al. 2015), consistent with the lack of enrichment for Nodal transcripts in the preimplantation mouse blastocyst.

At E6.5, the majority of human embryos treated with SB had lost NANOG expression, though they retained expression of SOX17 (Fig. 2.7B; (Blakeley et al. 2015)). Assessing whether the proportion of SOX17 cells remained unchanged was complicated by the lower levels of SOX17 detected in the majority of treated embryos (Blakeley et al. 2015). However, where SOX17 expression was detectable and quantifiable, the proportion of SOX17 cells was within the range of controls (Fig. 2.7C), suggesting that SB treatment destabilised NANOG expression in Epi cells but did not convert these cells to PrE. Therefore, this strongly suggests that Nodal signalling is essential for pluripotency in the human EPI. Curiously, previous experiments modulating TGFβ/Nodal signalling with SB instead observed a positive effect on NANOG expression and subsequent human ES cell derivation (Van der Jeught et al. 2013). However, a four-fold lower concentration of SB was used, and derivations were carried out on a supportive MEF layer (and thus in the presence of secreted Activin, (Greber et al. 2010), suggesting TGFβ/Nodal signalling may not be fully suppressed.

To test whether exogenous TGFβ/Nodal signalling could drive a lineage switch within the human embryo, similar to FGF signalling modulation in the mouse, blastocysts were treated with the TGFβ family ligand Activin A. Embryos were first treated with 10 ng/ml recombinant human Activin, a concentration commonly used to maintain human ES cell pluripotency (Vallier et al. 2005). E.5 blastocysts treated with 10 ng/ml Activin for 2 days expressed both NANOG and SOX17, in proportions similar to control embryos (Fig. 2.7B, 2.7C), suggesting that Activin may not be effective at this concentration *in vivo*.

Embryos were next treated with 50 ng/ml Activin, a concentration previously used to treat human embryos (Van der Jeught et al. 2013). Although the proportion of NANOG-expressing



D

Condition	Nanog	Nanog %	Sox17	Sox17 %	ICM	ICM %	Total cell no.
Control	8.5 ± 4.9	4.3 ± 3.2	5.75 ± 3.9	2.4 ± 0.8	14.5 ± 7.2	7.0 ± 3.6	205.5 ± 63.4
SB 40 μM D2 - D6/7	0.0	0.0	5.0 ± 0	3.7 ± 1.2	10.0 ± 7.1	8.3 ± 7.7	141.5 ± 46.0
Act 10 ng D5 - D6/7	6.5 ± 0.7	4.8 ± 2.4	5.5 ± 0.7	3.8 ± 1.1	14.0 ± 1.4	10.0 ± 3.1	149.5 ± 60.1
Act 50 ng D5 - D6/7	13.7 ± 5.1	4.9 ± 1.8	2.3 ± 3.2	1.1 ± 1.8	16.0 ± 2.0	6.1 ± 1.9	285.3 ± 0.3

Figure 2.7. Activin/Nodal signalling modulation in human embryos. (A) Treatment schedule. **(B)** Immunofluorence analysis for NANOG (green) and SOX17 (red) with DAPI merge (blue) at E6-7 in a control embryo (a) or in embryos treated with 40 μM SB-431542 from E2.5 (b), or ActivinA at 10 ng or 50 ng as indicated from E5 (c, d). Scale bars: 100 μm. Images in panels a and b taken by Norah Fogarty. 4 embryos were analysed as controls, 9 for SB, 2 for A 10 ng and 3 for A 50 ng. **(C)** Proportions of NANOG and SOX17-expressing cells in control or treated embryos as a percentage of total cell number as determined by MINS. Total ICM cell number also plotted. (**D**) Cell counts and calculated proportions used to generate graphs in (C). Control data is the same as that presented in the IGF1 treatment figure, 2 embryos (SB, A 10 ng) or 3 embryos (A 50 ng) were analysed for treatment experiments.

cells was again within range of controls, no SOX17 expression was observed in some embryos (Fig. 2.7B, 2.7C), suggesting Activin may instead be negatively affecting the PrE. Experiments are underway to confirm this, both increasing sample sizes under the conditions reported, but also treating embryos with Activin from earlier stages (E2.5) to determine if there is a greater effect, as was observed in the IGF1 study. Altogether, this shows that although Nodal signalling is required to maintain NANOG expression in the human Epi, endogenous levels are likely sufficient for this to occur, and further stimulation at the blastocyst stage does not affect the proportions of pluripotent cells. However, there is an indication that Activin/Nodal signalling may negatively affect gene expression in the PrE, and it will be interesting to determine if this holds true following further experiments.

2.2.3. Laminin-specific integrins are expressed in the human embryo

Initial human ES cell derivation and culture systems relied on a supportive MEF layer and the addition of serum or Knockout Serum Replacer as a means to preserve pluripotency *in vitro* (Thomson et al. 1998). However, the undefined nature of this culture system allowed for the presence of confounding factors that may instead antagonise the pluripotency support network. Supporting matrix substrate layers are commonly used in place of feeders, such as Matrigel, which consists mainly of laminin, though additional factors such as collagen and nidogen are also present (Kleinman et al. 1982). However, Matrigel is distilled from Engelbreth-Holm-Swarm (EHS) sarcoma cells and thus remains subject to batch-to-batch variability, and even in its growth factor-reduced form retains noticeable concentrations of various growth factors, including FGF (BD-Biosciences).

The basement membrane factors that constitute Matrigel, such as laminin and collagen, have also individually been used to support growth of human ES cells, and similar to others such as fibronectin and vitronectin, have also been included in various minimal media culture systems (Ludwig et al. 2006; Chen et al. 2011; Rodin et al. 2014; Takashima et al. 2014). Ideally, the

chosen basement membrane used *in vitro* would recapitulate binding profiles in the embryo, preferentially Epi-Epi to maintain the characteristics of the original niche, but also perhaps Epi-PE or Epi-TE. Once again, the human blastocyst gene expression dataset provides a valuable resource to investigate embryo-specific basement membrane components, at least at the transcript level, which could then be validated experimentally.

The RPKM boxplot dataset was used to explore the expression of integrin subunits, in order to determine whether any known integrin binding partners for commonly used basement membrane proteins (Humphries et al. 2006; Campbell and Humphries 2011) were enriched in the embryo. Transcripts for several integrins were identified (Fig 2.8A), with those representing the laminin-binding integrin α6β1 (*ITGA6, ITGB1*) enriched in both human ES cells and the blastocyst lineages (Fig 2.8A). Additional subunits *ITGA5* and *ITGAV* were detected in human ES cells, though expression fell below the significance threshold (Fig 2.8A). ITGA5 and ITGAV can complex with *ITGB1* as integrin α5β1and αVβ1receptors, which bind fibronectin and vitronectin respectively (Humphries et al. 2006; Campbell and Humphries 2011) This varied integrin profile may explain why existing human ES cells can be supported by multiple basement membranes. Previous studies have shown that integrin α6β1 preferentially binds laminin-111, laminin-511 and laminin-521 (Nishiuchi et al. 2006; Miyazaki et al. 2012; Nakagawa et al. 2014), and therefore the expression pattern of the transcripts associated with these laminins was also determined (Fig 2.8A). Transcripts for laminin-111 (*LAMA1, LAMB1, LAMC1*) were enriched across all blastocyst lineages and in human ES cells, as was the *LAMA5* subunit, though the latter fell just below the significance threshold (Fig 2.8A). Previous studies have demonstrated successful growth of human ES cells on laminins (Xu et al. 2001; Miyazaki et al. 2012), and laminin protein expression has been observed in both human ES cells and MEFs (Hongisto et al. 2012; Pook et al. 2015).

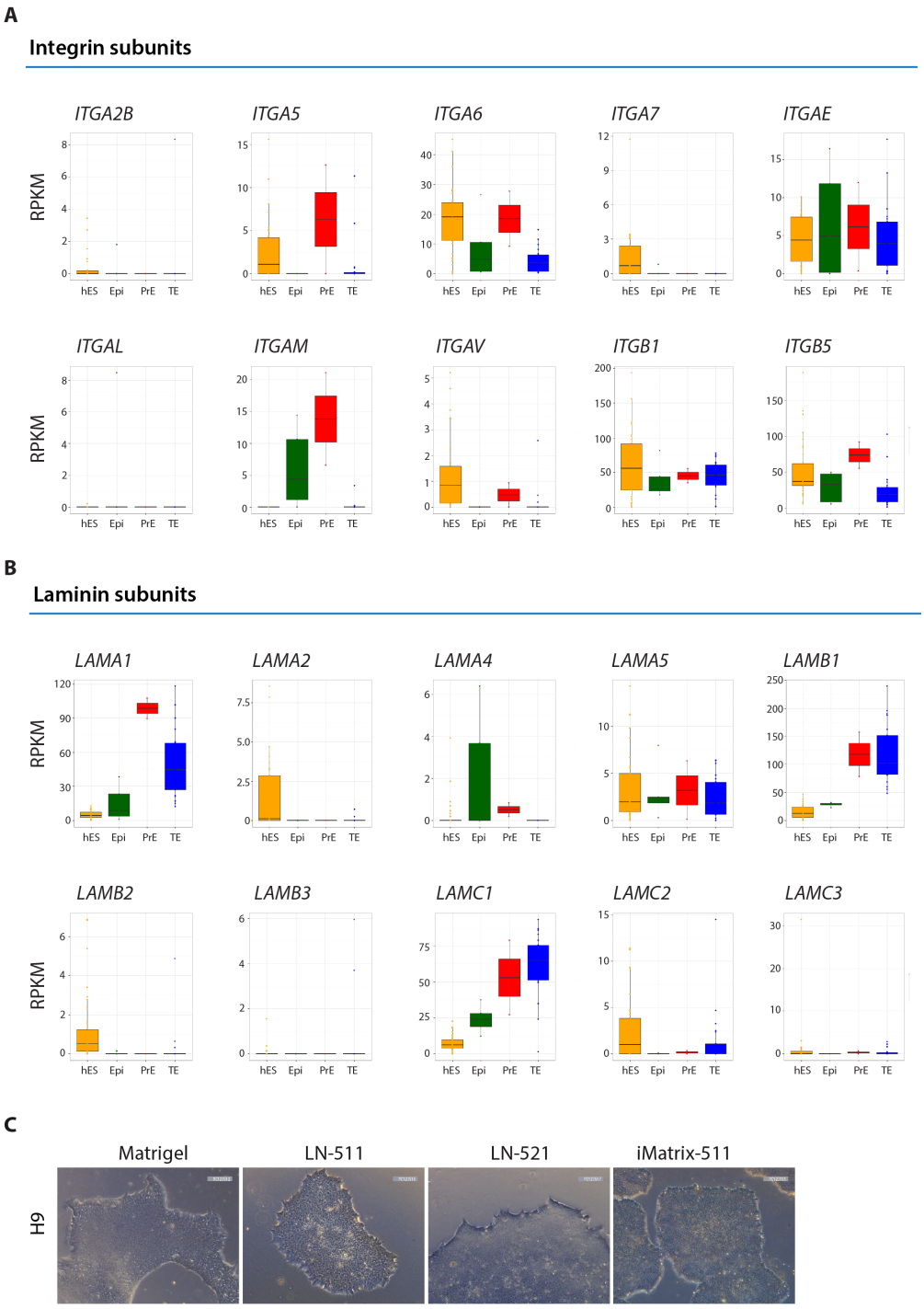


Figure 2.8. Laminin and associated integrin subunits are expressed in the Epi. Box plots of RPKM data showing the range of gene expression in human ES cells (yellow), or Epi (green), PrE (red) or TE (blue) blastocyst lineages based on gene expression patterns. Boxes correspond to the first and third quartiles, horizontal line to the median, whiskers extend to 1.5 times the interquartile range, and dots denote outliers. Plots are shown for **(A)** integrin subunits, and **(B)** laminin subunits, whose expression was detected in at least one blastocyst lineage or in ES cells. Box plots generated by Paul Blakeley. n = 23 hES, 5 Epi, 2 PE and 23 TE cells. **(C)** Representative images of H9 human ES cells grown on Matrigel or various commercially available laminins (LN-511, LN-521 from Biolamina; iMatrix-511 from Clontech/Takara).

Whole recombinant human laminin-511 (Biolamina, (Rodin et al. 2014)) and a recombinant Laminin-511 E8 fragment (iMatrix-511, Takara (Nakagawa et al. 2014)), were evaluated for their ability to maintain human ES cell pluripotency. Consistent with previous studies, both H1 and H9 human ES cell lines could be maintained on laminin in mTeSR media, as were maintained for multiple passages without losing ES cell morphology (Fig 2.8D). Although pluripotency gene expression is yet to be determined, previous work has demonstrated that laminin maintains OCT4 and NANOG gene expression in human ES cells, comparable to Matrigel (Takashima et al. 2014). Consequently, laminin appears to be a suitable embryo-specific basement membrane substrate capable of supporting existing human ES cells, and potentially deriving new ones. This is promising for future experiments that would seek to integrate signalling insights from the human blastocyst to design minimal chemically defined stem cell culture conditions that would better recapitulate the pluripotent niche.

2.3. Discussion

Although earlier studies indicated that inhibiting FGF/Erk signalling had no effect in the human embryo (Kuijk et al. 2012; Roode et al. 2012), they did not rule out the possibility of FGF functioning via an alternative downstream pathway as the FGF receptor was never targeted without also inhibiting Mek (Roode et al. 2012). This study finds that despite expression of FGF signalling pathway components in the human blastocyst, FGF stimulation negatively affects establishment of the pluripotent Epi and appears dispensable for maintenance. This is at odds with long-established human ES cell derivation and culture conditions, and challenges predominant assumptions about the signalling pathways regulating this lineage.

It is unclear why existing human ES cells are reliant on FGF signalling for their maintenance, as Activin/Nodal signalling seems to be the main pathway required to regulate pluripotency gene expression (Vallier et al. 2004; Vallier et al. 2005; Vallier et al. 2009). However, previous studies reported that treatment with the MEK inhibitor PD0325901 resulted in downregulation of

NANOG expression in human ES cells by 50% within 12 hours (Greber et al. 2010). Although treatment with SB-431542 resulted in greater NANOG downregulation over the same time period (85%), there were indications of a synergistic effect when both inhibitors were added (Greber et al. 2010), suggesting cooperative regulation of pluripotency by FGF and Activin/Nodal signalling, as previously reported (Vallier et al. 2005). As classical human ES cell derivation conditions involved addition of exogenous FGF, this may have promoted an *in vitro* adaptation of these cells towards FGF-dependency, perhaps as a substitute for an alternative ligand with similar downstream effects. Altogether, this further reinforces the finding that current human ES cells are distinct from the epiblast progenitors from which they are derived (Blakeley et al. 2015).

With this in mind, it seemed appropriate to take an unbiased approach and interrogate lineage-specific gene expression patterns in the human blastocyst directly. This study demonstrates that inhibiting TGFβ/Nodal signalling detrimentally affects NANOG expression, although supplementation with Activin does not seem to enhance the pluripotent compartment (though it may negatively affect gene expression in the PrE). This would seemingly suggest that endogenous NODAL signalling is sufficient to maintain the Epi *in vivo*, though further experiments treating human embryos with Activin from an earlier developmental stage, prior to epiblast maintenance, should shed further light on this.

How does this fit into the narrative of existing stem cell derivation and human ES cell maintenance, where Activin/Nodal signalling has been shown to regulate pluripotency gene expression? Perhaps exogenous Activin confers a greater benefit *in vitro* due to the lack of some other factor that would normally boost pluripotency, and whose absence is preventing human ES cells from secreting sufficient Nodal to support their continued self-renewal. Alternatively, a higher dose of Activin may be required to antagonise differentiation-promoting influences in culture, such as FGF for example, or BMP, which is present in serum (Ying et al. 2008), and has

been shown to destabilise pluripotency (Xu et al. 2002; Pera et al. 2004; Valera et al. 2010). In addition, although Nodal and Activin have been demonstrated to act equivalently in human ES cells, it is possible that Nodal may have a distinct effect in the human blastocyst. Activin transcripts (*INHBA* and *INHBB*) are not enriched in the human blastocyst, and given the distinctions emerging between human ES cells and the Epi compartment, it would be prudent to test the role of Nodal *in vivo* as well.

Remarkably, IGF signalling increases the proportion of both NANOG- and SOX17-expressing cells in the human ICM, suggesting this perhaps feeds into the proliferation and self-renewal circuitry of these lineages. Characterising global gene expression in IGF1-treated embryos and perhaps integrating this data into the single cell blastocyst dataset will be useful to determine the mechanics of this effect. A number of existing human ES culture media, supplements and basement membrane substrates contain IGF1 or insulin, including Knockout Serum Replacement (KOSR), N2B27, mTeSR and TeSR-E8, StemPro and Matrigel (BD-Biosciences ; Price et al. 1998; Ludwig et al. 2006; Wang et al. 2007; Chen et al. 2011; Ying and Smith 2012), suggesting this pathway may be unwittingly already supporting self-renewal. It would be interesting to determine if the proliferation effect in the embryo is specific to IGF1, or whether additional insulin/IGF1 signalling pathway ligands such as insulin or IGF2 may be capable of replicating this phenomenon.

Altogether, the transcriptomics dataset has been a useful window into human-specific gene expression. However, although initial steps are encouraging, with roles for insulin/IGF1 and TGFβ/Nodal signalling emerging with the human embryo, it still remains unclear what signals are responsible for specifying the Epi and PrE lineages within the human blastocyst. A caveat of the current analysis is that interrogating gene expression in the late blastocyst may be too late to identify signals required for specification, which may have been set in motion well before. Perhaps analysing the wider single cell dataset from earlier developmental stages may yield

results in this regard. Nevertheless, in the face of increasing divergence from both mouse embryonic developmental models and current human ES cells, the early human embryo represents a crucial system for investigating the fundamental biological mechanisms underlying cell fate choices.

Conclusions

The differences emerging between mouse and human gene expression in early development highlight the importance of further investigations in a human context. The human Epi seemingly does not depend on the same signalling pathways as the mouse – no reliance on LIF, no requirement for BMP/Id proteins to cooperate with STAT3 or antagonise FGF, which would otherwise promote differentiation. This raises interesting questions as to how lineage specification takes place in the human embryo, which this study attempted to address by investigating enrichment of signalling networks in single cells isolated at the blastocyst stage. One caveat of the dataset described here is that only one timepoint was explored, and at this stage cells could already be resolved into definitive lineages. Consequently, it is likely not possible to distinguish between lineage-specifying and lineage-maintenance networks, assuming that the mechanisms responsible for specification are set in motion earlier in development. However, this may be less of a concern if lineage segregation in the human does not follow the two-step mouse model of early ICM-TE segregation followed by later Epi-PE segregation within the ICM, as recent transcriptomic analysis suggests (Petropoulos et al. 2016). Cells in the early human embryo seem to retain developmental plasticity for longer than the mouse, though it is not clear why this is the case – one hypothesis is that the shorter developmental timeline in the mouse necessitates earlier lineage segregation (Rossant 2014), ensuring a fully committed TE is present in time for appropriate implantation to occur.

Nevertheless, the loss of NANOG expression observed here following TGFβ/Nodal signalling inhibition suggests this pathway feeds into the pluripotency gene regulatory network *in vivo*, as it does in existing human ES cells. Similarly, the ICM proliferation observed following insulin/IGF1 signalling activation suggests this pathway may feed into a self-renewal circuitry *in vivo* that is then perpetuated to allow derivation of ES cells, thereby capturing this otherwise transient population. Given that increased numbers of both NANOG- and SOX17-expressing

cells are observed following IGF1 treatment, it is unlikely to function via an Epi-specific mechanism, but could shed light on segregation of the ICM and the TE, as the TE appears unaffected. Although the current sample size is small, this hypothesis may be supported by the observation that IGF stimulation affects ICM proliferation when treated from the cleavage stage (E2), but not the early blastocyst stage (E5). Further IGF1 treatment experiments and similar quantification of ICM proportions will be crucial to support these initial findings, as well as investigating whether the expression of additional lineage-associated genes are also affected, including human Epi-specific markers such as KLF17.

A further point of contention is the intersection of lineage specifying genes with signalling networks during lineage establishment and maintenance in early development. This study finds that upregulating Gata6 in mouse ES cells enables FGF-independent upregulation of an ExEn gene program and cell fate conversion to iXEN cells, though previous mutant studies indicate that Gata6 and FGF likely converge to allow PrE specification in the mouse embryo. In future, mouse EpiSCs, which like the human Epi and ES cells depend on TGFβ/Nodal signalling (Brons et al. 2007; Tesar et al. 2007), would also represent an interesting system to further investigate the intersection of signalling networks and lineage-associated genes. Although FGF is routinely used to culture EpiSCs, the initial studies indicate that FGF-free support of EpiSCs is possible, as FGF mainly "improved the overall quality of the cultures, suggesting that it reinforces the efficiency of Activin signalling as in human ES cells" (Brons et al. 2007). Gata6 overexpression in EpiSCs, in the presence and absence of FGF, could therefore be a useful insight into the effect of ectopic Gata6 on pluripotency genes supported by Activin/Nodal signalling, with the added benefit that conditions for supporting mouse XEN cells are well defined, enabling an appropriate read out for conversion.

Given that GATA6 overexpression also upregulates a similar ExEn network in human ES cells, albeit incompletely, how does this correlate with the differing role of FGF in the mouse and

human embryo? Perhaps an alternative signalling pathway is required to enhance GATA6 function, and refining culture conditions to support putative human XEN cells could provide this. This however rests on the assumption that GATA6 plays a similar role in the human PrE specification hierarchy as it does in the mouse, which may not be the case given that GATA6 exhibits a broader expression pattern in the human blastocyst and is present in the TE (Roode et al. 2012; Deglincerti et al. 2016). Intriguingly, TE-associated genes *GATA3* and *CDX2* were also upregulated following *GATA6*-induction in human ES cells, which may indicate a broader spectrum of *GATA6* gene regulation, or simply be a by-product of pluripotency network downregulation.

A similar broad expression pattern has been observed for GATA6 in primate (Boroviak et al. 2015; Nakamura et al. 2016) and bovine embryos (Kuijk et al. 2012). However, this does not result in a consistent response to signalling, with GATA6 expression increased in response to FGF activation in bovine embryos but not completely ablated following inhibition (Kuijk et al. 2012), suggesting more than one regulatory mechanism is in place. In marmoset embryos, inhibition of FGF or WNT reduces the proportion of GATA6-exclusive expression, but increases colocalisation between GATA6 and CDX2 or NANOG, suggesting the main effect is blurring of lineage assignations (Boroviak et al. 2015). Conversely, FGF inhibition in human embryos has no effect on GATA6 expression (Kuijk et al. 2012; Roode et al. 2012) and GATA6 was still detected in blastocysts following FGF treatment in the study presented here. Further experiments should aim to quantify the proportion of GATA6 cells in FGF-treated embryos, as well as explore any changes in co-localisation dynamics to determine the broader effect on lineage-associated gene expression in our investigations.

It is clear that the human embryo represents an invaluable resource for study, and future comparisons with a variety of model organisms, especially with the increasing number of primate studies available, will no doubt vastly increase our understanding of the foundations of

early development. However, some challenges remain, such as availability of embryos donated to research, a relatively short time window for study, and difficulties with precision gene editing to investigate molecular hierarchy, although this latter point will likely shortly be overcome by CRISPR/Cas9 technology. Consequently, there remains a crucial requirement for an available stem cell population *in vitro* that is most reflective of the Epi counterpart *in vivo*. The demonstrable consistency in signalling requirements between mouse ES cells and the pre-implantation Epi (Boroviak et al. 2014) allows mouse ES cells to be used as a suitable model to test regulatory mechanisms. However, as existing human ES cells are somewhat distinct from the Epi both in gene expression and signalling requirements, this presents difficulties, especially in exploring the basic biology of this developmental stage.

It is likely that some of these issues can be overcome by using more suitable culture conditions. Several recent studies utilise a combination of inhibitors and growth factors in an attempt to generate human ES cells that are more similar to naïve mouse ES cells and the pluripotent cells of the human Epi (Chan et al. 2013; Gafni et al. 2013; Takashima et al. 2014; Theunissen et al. 2014; Valamehr et al. 2014; Ware et al. 2014; Guo et al. 2016). As discussed, some of these naïve cells successfully acquire characteristics of the human ICM, and as such will likely be a useful model for understanding the mechanisms involved in as global DNA hypomethylation or X-inactivation, for example. However, the resulting cell lines remain somewhat divergent from the human Epi, although some do upregulate human Epi-specific factors such as KLF17 (Blakeley et al. 2015; Guo et al. 2016; Collier et al. 2017). Furthermore, a number of these conditions still rely on addition of exogenous FGF or on FGF-secreting MEF support layers, and utilise inhibitors uncovered in a mouse embryo context that target pathways that may not be relevant in the human. With this in mind, it would be interesting to determine if the signalling insights uncovered in the current work and the future experiments proposed, could eventually inform an embryo-centric culture system that would permit derivation of human ES cells that more accurately reflect the pluripotent Epi compartment. Ideally these cells would exhibit a number of

the criteria discussed here, such as Epi-specific gene expression patterns and epigenetic profiles as determined by transcriptomic and methylomic analyses. Further tests could investigate if improving the starting culture conditions resulted in a broader differentiation potential (in contrast to existing human ES cell lines which exhibit biases towards particular lineages in directed differentiation assays (Osafune et al. 2008)), or resulted in improved rates of iPS cell derivation. These two latter points would likely be useful from a clinical as well as basic research context. In parallel to this, further characterisation of the effects of modulating these signalling pathways with growth factors or inhibitors *in vivo* may shed light on downstream signalling effectors and their relationship to lineage-associated genes, Altogether, this work presents an outline aimed at broadening our current understanding of early mouse and human embryonic development.

References

Abroun S, Ishikawa H, Tsuyama N, Liu S, Li FJ, Otsuyama K, Zheng X, Obata M, Kawano MM. 2004. Receptor synergy of interleukin-6 (IL-6) and insulin-like growth factor-I in myeloma cells that highly express IL-6 receptor alpha [corrected]. *Blood* **103**: 2291-2298.

Adachi T, Kar S, Wang M, Carr BI. 2002. Transient and sustained ERK phosphorylation and nuclear translocation in growth control. *Journal of cellular physiology* **192**: 151-159.

Adachi Y, Shibai Y, Mitsushita J, Shang WH, Hirose K, Kamata T. 2008. Oncogenic Ras upregulates NADPH oxidase 1 gene expression through MEK-ERK-dependent phosphorylation of GATA-6. *Oncogene* **27**: 4921 - 4932.

Aflatoonian B, Ruban L, Shamsuddin S, Baker D, Andrews P, Moore H. 2010. Generation of Sheffield (Shef) human embryonic stem cell lines using a microdrop culture system. *In Vitro Cellular & Developmental Biology - Animal* **46**: 236 - 241.

Amit M, Carpenter MK, Inokuma MS, Chiu C-P, Harris CP, Waknitz MA, Iskovitz-Eldor J, Thomson JA. 2000. Clonally Derived Human Embryonic Stem Cell Lines Maintain Pluripotency and Proliferative Potential for Prolonged Periods of Culture. *Developmental biology* **227**: 271 - 278.

Amit M, Shariki C, Margulets V, Iskovitz-Eldor J. 2004. Feeder Layer- and Serum-Free Culture of Human Embryonic Stem Cells1. *Biology of Reproduction* **70**: 837 - 845.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology* **11:R106**.

Arman E, Haffner-Krausz R, Chen Y, Heath JK, Lonai P. 1998. Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 5082-5087.

Artus J, Piliszek A, Hadjantonakis AK. 2011. The primitive endoderm lineage of the mouse blastocyst: sequential transcription factor activation and regulation of differentiation by Sox17. *Developmental biology* **350**: 393-404.

Avilion A, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & development* **17**: 126-140.

BD-Biosciences. Product Datasheet: BD Matrigel™ Basement Membrane Matrix Growth Factor Reduced Phenol Red Free (cat. no. 356231).

Beard C, Hochedlinger K, Plath K, Wutz A, Jaenisch R. 2006. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis (New York, NY : 2000)* **44**: 23-28.

Beattie GM, Lopez AD, Bucay N, Hinton A, Firpo MT, King CC, Hayek A. 2005. Activin A Maintains Pluripotency of Human Embryonic Stem Cells in the Absence of Feeder Layers. *Stem cells (Dayton, Ohio)* **23**: 489 - 495.

Beddington R, Robertson EJ. 1989. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**: 733 - 737.

Bendall SC, Stewart MH, Menendez P, George D, Vijayaragavan K, Werbowetski-Ogilvie T, Ramos-Mejia V, Rouleau A, Yang J, Bosse M et al. 2007. IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. *Nature* **448**: 1015-1021.

Besser D. 2004. Expression of Nodal, Lefty-A, and Lefty-B in Undifferentiated Human Embryonic Stem Cells Requires Activation of Smad2/3*. *The Journal of biological chemistry* **279**: 45076 - 45084.

Bessonnard S, De Mot L, Gonze D, Barriol M, Dennis C, Goldbeter A, Dupont G, Chazaud C. 2014. Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* **141**: 3637-3648.

Blair K, Wray J, Smith A. 2011. The liberation of embryonic stem cells. *PLoS Genetics* **7**: 2 - 6.

Blakeley P, Fogarty NM, Del Valle I, Wamaitha SE, Hu TX, Elder K, Snell P, Christie L, Robson P, Niakan KK. 2015. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**: 3613.

Bongso A, Fong C-Y, Ng S-C, Ratnam S. 1994. Isolation and culture of inner cell mass cells from human blastocysts. *Human Reproduction* **9**: 2110 - 2117.

Boroviak T, Loos R, Bertone P, Smith A, Nichols J. 2014. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nature Cell Biology* **16**: 516-528.

Boroviak T, Loos R, Lombard P, Okahara J, Behr R, Sasaki E, Nichols J, Smith A, Bertone P. 2015. Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Developmental cell* **35**: 366-382.

Bourillot PY, Aksoy I, Schreiber V, Wianny F, Schulz H, Hummel O, Hubner N, Savatier P. 2009. Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog. *Stem cells (Dayton, Ohio)* **27**: 1760-1771.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**: 947-956.

Bradley A, Evans MJ, Kaufman M, Robertson EJ. 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**: 255 - 256.

Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK, Clarkson A, Ahrlund-Richter L, Pedersen RA et al. 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**: 191-195.

Brown S, Teo A, Pauklin S, Hannan N, Cho CH, Lim B, Vardy L, Dunn NR, Trotter MW, Pedersen RA et al. 2011. Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem cells (Dayton, Ohio)* **29**: 1176 - 1185.

Brunet A, Roux D, Lenormand P, Dowd S, Keyse S, Pouyssegur J. 1999. Nuclear translocation of p42/p44 mitogen-activated protein kinase is required for growth factor-induced gene expression and cell cycle entry. *The EMBO journal* **18**: 664-674.

Buehr M, Meek S, Blair K, Yang J, Ure J, Silva J, McLay R, Hall J, Ying QL, Smith A. 2008. Capture of authentic embryonic stem cells from rat blastocysts. *Cell* **135**: 1287-1298.

Burdon T, Stracey C, Chambers I, Nichols J, Smith A. 1999. Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Developmental biology* **210**: 30-43.

Callahan JF, Burgess JL, Fornwald JA, Gaster LM, Harling JD, Harrington FP, Heer J, Kwon C, Lehr R, Mathur A et al. 2002. Identification of Novel Inhibitors of the Transforming Growth Factor β1 (TGF-β1) Type 1 Receptor (ALK5). *Journal of Medicinal Chemistry* **45**: 999-1001.

Campbell IA, Humphries MJ. 2011. Integrin Structure, Activation, and Interactions. *Cold Spring Harbor Perspectives in Biology* **3:a004994**.

Capo-Chichi CD, Rula ME, Smedberg JL, Vanderveer L, Parmacek MS, Morrisey EE, Godwin AK, Xu XX. 2005. Perception of differentiation cues by GATA factors in primitive endoderm lineage determination of mouse embryonic stem cells. *Developmental biology* **286**: 574-586.

Cauffman G, De Rycke M, Sermon K, Liebaers I, Van de Velde H. 2009. Markers that define stemness in ESC are unable to identify the totipotent cells in human preimplantation embryos. *Human Reproduction* **24**: 63-70.

Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedle S, Smith A. Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell* **113**: 643 - 655.

Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, Vrana J, Jones K, Grotewold L, Smith A. 2007. Nanog safeguards pluripotency and mediates germline development. *Nature* **450**: 1230-1234.

Chan YS, Goke J, Lu X, Venkatesan N, Feng B, Su IH, Ng HH. 2013. A PRC2-dependent repressive role of PRDM14 in human embryonic stem cells and induced pluripotent stem cell reprogramming. *Stem cells (Dayton, Ohio)* **31**: 682-692.

Chazaud C, Yamanaka Y. 2016. Lineage specification in the mouse preimplantation embryo. *Development* **143**: 1063-1074.

Chazaud C, Yamanaka Y, Pawson T, Rossant J. 2006. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Developmental cell* **10**: 615-624.

Chen G, Gulbranson DR, Hou Z, Bolin JM, Ruotti V, Probasco MD, Smuga-Otto K, Howden SE, Diol NR, Propson NE et al. 2011. Chemically defined conditions for human iPSC derivation and culture. *Nature Methods* **8**: 424 - 429.

Chen T, Dent SY. 2014. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nature Reviews Genetics* **15**: 93-106.

Chen WS, Manova K, Wieinstein DC, Duncan SA, Plump AS, Prezioso VR, Bacharova RF, Darnell JEJ. Disruption of the HNF-4 gene, expressed in visceral endoderm, leads to cell death in embryonic ectoderm and impaired gastrulation of mouse embryos. *Genes & development* **8**: 2466 - 2477.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106-1117.

Cheng AM, Saxton TM, Sakai R, Kulkarni S, Mbamalu G, Vogel W, Tortorice CG, Cardiff RD, Cross JC, Muller WJ et al. 1998. Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. *Cell* **95**: 793-803.

Cho LT, Wamaitha SE, Tsai IJ, Artus J, Sherwood RI, Pedersen RA, Hadjantonakis AK, Niakan KK. 2012. Conversion from mouse embryonic to extra-embryonic endoderm stem cells reveals distinct differentiation capacities of pluripotent stem cell states. *Development* **139**: 2866-2877.

Choi Hyun W, Joo Jin Y, Hong Yean J, Kim Jong S, Song H, Lee Jeong W, Wu G, Schöler Hans R, Do Jeong T. 2016. Distinct Enhancer Activity of Oct4 in Naive and Primed Mouse Pluripotency. *Stem cell reports* **7**: 911-926.

Collier AJ, Panula SP, Schell JP, Chovanec P, Plaza Reyes A, Petropoulos S, Corcoran AE, Walker R, Douagi I, Lanner F et al. 2017. Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. *Cell stem cell* **20**: 874-890 e877.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, NY)* **339**: 819 - 823.

Conti L, Pollard SM, Gorba T, Reitano E, Toselli M, Biella G, Sun Y, Sanzone S, Ying QL, Cattaneo E et al. 2005. Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biology* **3**: e283.

Costa M, Marchi M, Cardarelli F, Roy A, Beltram F, Maffei L, Ratto GM. 2006. Dynamic regulation of ERK2 nuclear translocation and mobility in living cells. *Journal of cell science* **119**: 4952-4963.

Cowan C, Klimanskaya I, McMahon J, Atienza J, Witmeyer J, Zucker JP, Wang S, Morton CC, McMahon AP, Powers D et al. 2004. Derivation of embryonic stem-cell lines from human blastocysts. *New England Journal of Medicine* **350**: 1353 - 1356.

Daheron L, Opitz SL, Zaehres H, Lensch WM, Andrews PW, Iskovitz-Eldor J, Daley GQ. 2004. LIF/STAT3 signalling fails to maintain self-renewal of human embryonic stem cells. *Stem cells (Dayton, Ohio)* **22**: 770 - 778.

Davidson KC, Adams AM, Goodson JM, McDonald CE, Potter JC, Berndt JD, Biechele TL, Taylor RJ, Moon RT. 2012. Wnt/β-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 4485 - 4490.

De Los Angeles A, Ferrari F, Xi R, Fujiwara Y, Benvenisty N, Deng H, Hochedlinger K, Jaenisch R, Lee S, Leitch HG et al. 2015. Hallmarks of pluripotency. *Nature* **525**: 469-478.

De Paepe C, Cauffman G, Verloes A, Sterckx J, Devroey P, Tournaye H, Liebaers I, Van de Velde H. 2013. Human trophectoderm cells are not yet committed. *Human Reproduction* **28**: 740-749.

De Paepe C, Krivega M, Cauffman G, Geens M, Van de Velde H. 2014. Totipotency and lineage segregation in the human embryo. *Molecular Human Reproduction* **20**: 599-618.

Deglincerti A, Croft GF, Pietila LN, Zernicka-Goetz M, Siggia ED, Brivanlou AH. 2016. Self-organization of the in vitro attached human embryo. *Nature* **533**: 251-254.

Dietrich JE, Hiiragi T. 2007. Stochastic patterning in the mouse pre-implantation embryo. *Development* **134**: 4219-4231.

Dvorak P, Dvorakova D, Koskova S, Vodinska M, Najvirtova M, Krekac D, Hampl A. 2005. Expression and potential role of fibroblast growth factor 2 and its receptors in human embryonic stem cells. *Stem cells (Dayton, Ohio)* **23**: 1200-1211.

Eden E, Lipson D, Yogev S, Yakhini Z. 2007. Discovering Motifs in Ranked Lists of DNA sequences. *PLoS Computational Biology* **3(3)**.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: A Tool For Discovery And Visualization of Enriched GO Terms in Ranked Gene Lists. *BMC Bioinformatics* **10**.

Evans MJ, Kaufman MH. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**: 154-156.

Feldman B, Poueymirou W, Papaioannou VE, DeChiara TM, Goldfarb M. 1995. Requirement of FGF-4 for postimplantation mouse development. *Science (New York, NY)* **267**: 246-249.

Festuccia N, Osorno R, Halbritter F, Karwacki-Neisius V, Navarro P, Colby D, Wong F, Yates A, Tomlinson SR, Chambers I. 2012. Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell stem cell* **11**: 477-490.

Findlay GM, Smith MJ, Lanner F, Hsiung MS, Gish GD, Petsalaki E, Cockburn K, Kaneko T, Huang H, Bagshaw RD et al. 2013. Interaction domains of Sos1/Grb2 are finely tuned for cooperative control of embryonic stem cell fate. *Cell* **152**: 1008-1020.

Frankenberg S, Gerbe F, Bessonnard S, Belville C, Pouchin P, Bardot O, Chazaud C. 2011. Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Developmental cell* **21**: 1005-1013.

Frum T, Halbisen MA, Wang C, Amiri H, Robson P, Ralston A. 2013. Oct4 cell-autonomously promotes primitive endoderm development in the mouse blastocyst. *Developmental cell* **25**: 610-622.

Fujikura J, Yamato E, Yonemura S, Hosoda K, Masui S, Nakao K, Miyazaki Ji J, Niwa H. 2002. Differentiation of embryonic stem cells is induced by GATA factors. *Genes & development* **16**: 784-789.

Gafni O, Weinberger L, Mansour AA, Manor YS, Chomsky E, Ben-Yosef D, Kalma Y, Viukov S, Maza I, Zviran A et al. 2013. Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**: 282-286.

Gardner DK, Schoolcraft WB. 1999. In vitro culture of human blastocyst. in *Towards reproductive certainty* (eds. R Jansen, D Mortimer), pp. 378 - 388. Parthenon Press, Carnforth.

Gardner RL. 1985. Clonal analysis of early mammalian development. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **312**: 163-178.

Gardner RL, Brook FA. 1997. Reflections on the biology of embryonic stem (ES) cells. *The International Journal of Developmental Biology* **41**: 235 - 243.

Gardner RL, Papaioannou VE, Barton SC. 1973. Origin of the ectoplacental cone and secondary giant cells in mouse blastocysts reconstituted from isolated trophoblast and inner cell mass. *Journal of embryology and experimental morphology* **30**: 561-572.

Gardner RL, Rossant J. 1979. Investigation of the fate of 4-5 day post-coitum mouse inner cell mass cells by blastocyst injection. *Journal of embryology and experimental morphology* **52**: 141-152.

Gautier A, Deiters A, Chin JW. 2011. Light-activated kinases enable temporal dissection of signaling networks in living cells. *Journal of the American Chemical Society* **133**: 2124-2127.

Gerbe F, Cox B, Rossant J, Chazaud C. 2008. Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst. *Developmental biology* **313**: 594-602.

Giroux S, Tremblay M, Bernard D, Cardin-Girard JF, Aubry S, Larouche L, Rousseau S, Huot J, Landry J, Jeannotte L et al. 1999. Embryonic death of Mek1-deficient mice reveals a role for this kinase in angiogenesis in the labyrinthine region of the placenta. *Current biology : CB* **9**: 369-372.

Goldin SN, Papaioannou VE. 2003. Paracrine action of FGF4 during periimplantation development maintains trophectoderm and primitive endoderm. *Genesis (New York, NY : 2000)* **36**: 40-47.

Grabarek JB, Zyzynska K, Saiz N, Piliszek A, Frankenberg S, Nichols J, Hadjantonakis AK, Plusa B. 2012. Differential plasticity of epiblast and primitive endoderm precursors within the ICM of the early mouse embryo. *Development* **139**: 129-139.

Granier C, Gurchenkov V, Perea-Gomez A, Camus A, Ott S, Papanayotou C, Iranzo J, Moreau A, Reid J, Koentges G et al. 2011. Nodal cis-regulatory elements reveal epiblast and primitive endoderm heterogeneity in the peri-implantation mouse embryo. *Developmental biology* **349**: 350-362.

Greber B, Coulon P, Zhang M, Moritz S, Frank S, Muller-Molina AJ, Arauzo-Bravo MJ, Han DW, Pape HC, Scholer HR. 2011. FGF signalling inhibits neural induction in human embryonic stem cells. *The EMBO journal* **30**: 4874-4884.

Greber B, Lehrach H, Adjaye J. 2007. Fibroblast growth factor 2 modulates transforming growth factor beta signaling in mouse embryonic fibroblasts and human ESCs (hESCs) to support hESC self-renewal. *Stem cells (Dayton, Ohio)* **25**: 455-464.

Greber B, Wu G, Bernemann C, Joo JY, Han DW, Ko K, Tapia N, Sabour D, Sterneckert J, Tesar P et al. 2010. Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. *Cell stem cell* **6**: 215-226.

Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND, Robson P. 2010. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Developmental cell* **18**: 675-685.

Guo G, von Meyenn F, Santos F, Chen Y, Reik W, Bertone P, Smith A, Nichols J. 2016. Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem cell reports* **6**: 437 - 446.

Guo G, Yang J, Nichols J, Hall JS, Eyres I, Mansfield W, Smith A. 2009. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**: 1063-1069.

Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, Yan J, Ren X, Lin S, Li J et al. 2014. The DNA methylation landscape of human early embryos. *Nature* **511**: 606-610.

Hadjantonakis AK, Gertsenstein M, Ikawa M, Okabe M, Nagy A. 1998. Generating green fluorescent mice by germline transmission of green fluorescent ES cells. *Mechanisms of Development* **76**: 79 - 90.

Hall J, Guo G, Wray J, Eyres I, Nichols J, Grotewold L, Morfopoulou S, Humphreys P, Mansfield W, Walker R et al. 2009. Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. *Cell stem cell* **5**: 597-609.

Hamatani T. CMG, Sharov A.A., Ko M.S.H. 2004. Dynamics of global gene expression changes during mouse preimplantation development. *Developmental cell* **6**: 117-131.

Hamazaki T, Kehoe SM, Nakano T, Terada N. 2006. The Grb2/Mek pathway represses Nanog in murine embryonic stem cells. *Molecular and cellular biology* **26**: 7539-7549.

Hamilton WB, Brickman JM. 2014. Erk signaling suppresses embryonic stem cell self-renewal to specify endoderm. *Cell Reports* **9**: 2056-2070.

Hamilton WB, Kaji K, Kunath T. 2013. ERK2 suppresses self-renewal capacity of embryonic stem cells, but is not required for multi-lineage commitment. *PloS one* **8**: e60907.

Han DW, Greber B, Wu G, Tapia N, Arauzo-Bravo MJ, Ko K, Bernemann C, Stehling M, Scholer HR. 2011. Direct reprogramming of fibroblasts into epiblast stem cells. *Nature Cell Biology* **13**: 66 - 71.

Han DW, Tapia N, Joo JY, Greber B, Arauzo-Bravo MJ, Bernemann C, Ko K, Wu G, Stehling M, Do JT et al. 2010. Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. *Cell* **143**: 617-627.

Hanna JH, Cheng AW, Saha K, Kim J, Lengner CJ, Soldner F, Cassady JP, Muffat J, Carey BW, Jaenisch R. 2010. Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 9222 - 9227.

Harvey MB, Kaye PL. 1992. Insulin-like growth factor-1 stimulates growth of mouse preimplantation embryos in vitro. *Molecular Reproduction and Development* **31**: 195 -199.

Hatano N, Mori Y, Oh-hora M, Kosugi A, Fujikawa T, Nakai N, Niwa H, Miyazaki J, Hamaoka T, Ogata M. 2003. Essential role for ERK2 mitogen-activated protein kinase in placental development. *Genes to cells : devoted to molecular & cellular mechanisms* **8**: 847-856.

Hebenstreit D, Miaoqing F, Muxin G, Charoensawan V, van Oudenaarden A, Teichmann SA. 2011. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Molecular systems biology* **7**: 3 - 9.

Hochedlinger K, Yamada Y, Beard C, Jaenisch R. 2005. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **121**: 465-477.

Hongisto H, Vuoristo S, Mikhailova A, Suuronen R, Virtanen I, Otonkoski T, Skottman H. 2012. Laminin-511 expression is associated with the functionality of feeder cells in human embryonic stem cell culture. *Stem Cell Research* **8**: 97-108.

Huang K, Maruyama T, Fan G. 2014. The Naive State of Human Pluripotent Stem Cells: A Synthesis of Stem Cell and Preimplantation Embryo Transcriptome Analyses. *Cell stem cell* **15**: 410-415.

Huang Y, Osorno R, Tsakiridis A, Wilson V. 2012. In Vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. *Cell Reports* **2**: 1571 - 1578.

Humphrey RK, Beattie GM, Lopez AD, Bucay N, King CC, Firpo MT, Rose-John S, Hayek A. 2004. Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. *Stem cells (Dayton, Ohio)* **22**: 522-530.

Humphries JD, Byron A, Humphries MJ. 2006. Integrin ligands at a glance. *Journal of cell science* **119**: 3901 - 3903.

Inman GJ, Nicolas FJ, Callahan JF, Harling JD, Gaster LM, Reith AD, Laping NJ, Hill CS. 2002. SB-431542 Is a Potent and Specific Inhibitor of Transforming Growth Factor- Superfamily Type I Activin Receptor-Like Kinase (ALK) Receptors ALK4, ALK5, and ALK7. *Molecular Pharmacology* **62**: 65 - 74.

Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y, Lemischka IR. 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**: 533 - 538.

James D, Levine AJ, Besser D, Brivanlou AH. 2005. TGFβ/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132**: 1273 - 1282.

Kaestner KH, Hiemisch H, Schutz G. 1998. Targeted disruption of the gene encoding hepatocyte nuclear factor 3gamma results in reduced transcription of hepatocyte-specific genes. *Molecular and cellular biology* **18**: 4245 - 4251.

Kanai-Azuma M, Kanai Y, Gad JM, Tajima Y, Taya C, Kurohmaru M, Sanai Y, Yonekawa H, Yazaki K, Tam PP et al. 2002. Depletion of definitive gut endoderm in Sox17-null mice. *Development* **129**.

Kang M, Piliszek A, Artus J, Hadjantonakis AK. 2013. FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. *Development* **140**: 267-279.

Kim MO, Kim SH, Cho YY, Nadas J, Jeong CH, Yao K, Kim DJ, Yu DH, Keum YS, Lee KY et al. 2012. ERK1 and ERK2 regulate embryonic stem cell self-renewal through phosphorylation of Klf4. *Nature structural & molecular biology* **19**: 283-290.

Kimura-Yoshida C, Tian E, Nakano H, Amazaki S, Shimokawa K, Rossant J, Aizawa S, Matsuo I. 2007. Crucial roles of Foxa2 in mouse anterior-posterior axis polarization via regulation of anterior visceral endoderm-specific genes. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 5919 - 5924.

Kinoshita M, Shimosato D, Yamane M, Niwa H. 2015. Sox7 is dispensable for primitive endoderm differentiation from mouse ES cells. *BMC developmental biology* **15**: 37.

Kleinman HK, McGarvey ML, Liotta LA, Robey PG, Tryggvason K, Martin GR. 1982. Isolation and characterization of type IV procollagen, laminin, and heparan sulfate proteoglycan from the EHS sarcoma. *Biochemistry* **21**: 6188 - 6193.

Kleinsmith LJ, Pierce GBJ. 1964. Multipotentiality of single embryonal carcinoma cells. *Cancer Research* **24**: 1544-1551.

Kobold S, Guhr A, Kurtz A, Loser P. 2015. Human embryonic and induced pluripotent stem cell research trends: complementation and diversification of the field. *Stem cell reports* **4**: 914-925.

Kojima Y, Kaufman-Francis K, Studdert Joshua B, Steiner Kirsten A, Power Melinda D, Loebel David AF, Jones V, Hor A, de Alencastro G, Logan Grant J et al. 2014a. The Transcriptional and Functional Properties of Mouse Epiblast Stem Cells Resemble the Anterior Primitive Streak. *Cell stem cell* **14**: 107-120.

Kojima Y, Tam OH, Tam PPL. 2014b. Timing of developmental events in the early mouse embryo. *Seminars in Cell & Developmental Biology* **34**: 65-75.

Koutsourakis M, Langeveld A, Patient R, Beddington R, Grosveld F. 1999. The transcription factor GATA6 is essential for early extraembryonic development. *Development* **126**: 723-732.

Kuijk EW, van Tol LT, Van de Velde H, Wubbolts R, Welling M, Geijsen N, Roelen BA. 2012. The roles of FGF and MAP kinase signaling in the segregation of the epiblast and hypoblast cell lineages in bovine and human embryos. *Development* **139**: 871-882.

Kunath T, Arnaud D, Uy GD, Okamoto I, Chureau C, Yamanaka Y, Heard E, Gardner RL, Avner P, Rossant J. 2005. Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. *Development* **132**: 1649-1661.

Kunath T, Saba-El-Leil MK, Almousailleakh M, Wray J, Meloche S, Smith A. 2007. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**: 2895-2902.

Kunath T, Yamanaka Y, Detmar J, MacPhee D, Caniggia I, Rossant J, Jurisicova A. 2014. Developmental differences in the expression of FGF receptors between human and mouse embryos. *Placenta* **35**: 1079-1088.

Kurimoto K, Yabuta Y, Ohinata Y, Ono Y, Uno KD, Yamada RG, Ueda HR, Saitou M. 2006. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Research* **34**: e42.

Lagalla C, Barberi M, Orlando G, Sciajno R, Bonu MA, Borini A. 2015. A quantitative approach to blastocyst quality evaluation: morphometric analysis and related IVF outcomes. *Journal of Assisted Reproduction and Genetics* **32**: 705 - 712.

Lanner F, Rossant J. 2010. The role of FGF/Erk signaling in pluripotent cells. *Development* **137**: 3351-3360.

Le Bin GC, Munoz-Descalzo S, Kurowski A, Leitch H, Lou X, Mansfield W, Etienne-Dumeau C, Grabole N, Mulas C, Niwa H et al. 2014. Oct4 is required for lineage priming in the developing inner cell mass of the mouse blastocyst. *Development* **141**: 1001-1010.

Leitch HG, McEwen KR, Turp A, Encheva V, Carroll T, Grabole N, Mansfield W, Nashun B, Knezovich JG, Smith A et al. 2013. Naive pluripotency is associated with global DNA hypomethylation. *Nature structural & molecular biology* **20**: 311-316.

Levenstein ME, Ludwig TE, Xu RH, Llanas RA, VanDenHeuvel-Kramer K, Manning D, Thomson JA. 2006. Basic fibroblast growth factor support of human embryonic stem cell self-renewal. *Stem cells (Dayton, Ohio)* **24**: 568-574.

Li G, Ren C, Shi J, Huang W, Liu H, Feng X, Liu W, Zhu B, Zhang C, Wang L et al. 2013. Identification, expression and subcellular localization of ESRG. *Biochemical and Biophysical Research Communications* **435**: 160 - 164.

Li L, Arman E, Ekblom P, Edgar D, Murray P, Lonai P. 2004. Distinct GATA6- and laminin-dependent mechanisms regulate endodermal and ectodermal embryonic stem cell fates. *Development* **131**: 5277-5286.

Lighten AD, Hardy K, Winston RML, Moore GE. 1997. Expression of mRNA for the insulin-like growth factors and their receptors in human preimplantation embryos. *Molecular Reproduction and Development* **47**: 134 - 139.

Lighten AD, Moore GE, Winston RML, Hardy K. 1998. Routine addition of human insulin-like growth factor-I ligand could benefit clinical in-vitro fertilization culture. *Human Reproduction* **13**: 3144 - 3150.

Lim CY, Tam WL, Zhang J, Ang HS, Jia H, Lipovich L, Ng HH, Wei CL, Sung WK, Robson P et al. 2008. Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages. *Cell stem cell* **3**: 543 - 554.

Loh KM, Ang LT, Zhang J, Kumar V, Ang J, Auyeong JQ, Lee KL, Choo SH, Lim CY, Nichane M et al. 2014. Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell stem cell* **14**: 237-252.

Loh KM, Lim B, Ang LT. 2015. Ex uno plures: molecular designs for embryonic pluripotency. *Physiological Reviews* **95**: 245-295.

Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J et al. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature genetics* **38**: 431-440.

Loser P, Schirm J, Guhr A, Wobus AM, Kurtz A. 2010. Human embryonic stem cell lines and their use in international research. *Stem cells (Dayton, Ohio)* **28**: 240-246.

Lou X, Kang M, Xenopoulos P, Munoz-Descalzo S, Hadjantonakis AK. 2014. A rapid and efficient 2D/3D nuclear segmentation method for analysis of early mouse embryo and stem cell image data. *Stem cell reports* **2**: 382-397.

Ludwig TE, Levenstein ME, Jones JM, Berggren WT, Mitchen ER, Frane JL, Crandall LJ, Daigh CA, Conard KR, Piekarczyk MS et al. 2006. Derivation of human embryonic stem cells in defined conditions. *Nature biotechnology* **24**: 185-187.

Lyashenko N, Winter M, Migliorini D, Biechele T, Moon RT, Hartmann C. 2011. Differential requirement for the dual functions of β-catenin in embryonic stem cell self-renewal and germ layer formation. *Nature Cell Biology* **13**: 753 - 761.

Ma Z, Swigut T, Valouev A, Rada-Iglesias A, Wysocka J. 2011. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nature structural & molecular biology* **18**: 120-127.

Malaguarnera R, Belfiore A. 2014. The emerging role of insulin and insulin-like growth factor signaling in cancer stem cells. *Frontiers in Endocrinology* **5**: 10.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science (New York, NY)* **339**: 823 - 826.

Martello G, Bertone P, Smith A. 2013. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *The EMBO journal* **32**: 2561-2574.

Martin GR. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **78**: 7634-7638.

Martin GR, Evans MJ. 1974. The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. *Cell* **2**: 163 - 172.

Masaki H, Kato-Itoh M, Umino A, Sato H, Hamanaka S, Kobayashi T, Yamaguchi T, Nishimura K, Ohtaka M, Nakanishi M et al. 2015. Interspecific in vitro assay for the chimera-forming ability of human pluripotent stem cells. *Development* **142**: 3222-3230.

Mascetti Victoria L, Pedersen Roger A. 2016. Human-Mouse Chimerism Validates Human Stem Cell Pluripotency. *Cell stem cell* **18**: 67-72.

Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA et al. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology* **9**: 625-635.

McDonald AC, Biechele S, Rossant J, Stanford WL. 2014. Sox17-mediated XEN cell conversion identifies dynamic networks controlling cell-fate decisions in embryo-derived stem cells. *Cell Reports* **9**: 780-793.

Mitalipova M, Calhoun J, Shin S, Wininger D, Schulz T, Noggle S, Venable A, Lyons I, Robins A, Stice S. 2003. Human embryonic stem cell lines derived from discarded embryos. *Stem cells (Dayton, Ohio)* **21**: 521 - 526.

Mitsui K, Tokuzawa Y, Itoh N, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S. 2003. The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* **113**: 631 - 642.

Miyazaki T, Futaki S, Suemori H, Taniguchi Y, Yamada M, Kawasaki M, Hayashi M, Kumagai H, Nakatsuji N, Sekiguchi K et al. 2012. Laminin E8 fragments support efficient adhesion and expansion of dissociated human pluripotent stem cells. *Nature Communications* **3**: 1236.

Molkentin JD, Lin Q, Duncan SA, Olson EN. 1997. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes & development* **11**: 1061-1072.

Morrisey EE, Musco S, Chen MY, Lu MM, Leiden JM, Parmacek MS. 2000. The gene encoding the mitogen-responsive phosphoprotein Dab2 is differentially regulated by GATA-6 and GATA-4 in the visceral endoderm. *The Journal of biological chemistry* **275**: 19949-19954.

Morrisey EE, Tang Z, Sigrist K, Lu MM, Jiang F, Ip HS, Parmacek MS. 1998. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes & development* **12**: 3579-3590.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**: 621 - 628.

Nagy A, Rossant J, Nagy R, Abramow-Newerly W, Roder JC. 1993. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 8424 - 8428.

Najm FJ, Chenoweth JG, Anderson PD, Nadeau JH, Redline RW, McKay RDG, Tesar PJ. 2011. Isolation of epiblast stem cells from preimplantation mouse embryos. *Cell stem cell* **8**: 318-325.

Nakagawa M, Taniguchi Y, Senda S, Takizawa N, Ichisaka T, Asano K, Morizane A, Doi D, Takahashi J, Nishizawa M et al. 2014. A novel efficient feeder-free culture system for the derivation of human induced pluripotent stem cells. *Scientific Reports* **4**: 3594.

Nakamura T, Okamoto I, Sasaki K, Yabuta Y, Iwatani C, Tsuchiya H, Seita Y, Nakamura S, Yamamoto T, Saitou M. 2016. A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**: 57-62.

Niakan KK, Eggan K. 2013. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Developmental biology* **375**: 54-64.

Niakan KK, Ji H, Maehr R, Vokes SA, Rodolfa KT, Sherwood RI, Yamaki M, Dimos JT, Chen AE, Melton DA et al. 2010. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes & development* **24**: 312-326.

Nichols J, Branko Z, Anastassiadis K, Niwa H, Kiewe-Nebenius D, Chambers I, Scholer HR, Smith A. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**: 379-391.

Nichols J, Evans EP, Smith A. 1990. Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development* **110**: 1341 - 1348.

Nichols J, Silva J, Roode M, Smith A. 2009. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* **136**: 3215-3222.

Nishiuchi R, Takagi J, Hayashi M, Ido H, Yagi Y, Sanzen N, Tsuji T, Yamada M, Sekiguchi K. 2006. Ligand-binding specificities of laminin-binding integrins: a comprehensive survey of laminin-integrin interactions using recombinant alpha3beta1, alpha6beta1, alpha7beta1 and alpha6beta4 integrins. *Matrix Biology* **25**: 189-197.

Niwa H, Burdon T, Chambers I, Smith A. 1998. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes & development* **12**: 2048-2060.

Niwa H, Miyazaki J, Smith A. 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature genetics* **24**: 372 - 376.

Niwa H, Ogawa K, Shimosato D, Adachi K. 2009. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460**: 118-122.

Nordhoff V, Hubner K Fau - Bauer A, Bauer A Fau - Orlova I, Orlova I Fau - Malapetsa A, Malapetsa A Fau - Scholer HR, Scholer HR. 2001. Comparative analysis of human, bovine, and murine Oct-4 upstream promoter sequences.

Ornitz DM, Itoh N. 2015. The Fibroblast Growth Factor signaling pathway. *Wiley Interdiscip Rev Dev Biol* **4**: 215-266.

Osafune K, Caron L, Borowiak M, Martinez RJ, Fitz-Gerald CS, Sato Y, Cowan CA, Chien KR, Melton DA. 2008. Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat Biotech* **26**: 313-315.

Pages G, Guerin S, Grall D, Bonino F, Smith A, Anjuere F, Auberger P, Pouyssegur J. 1999. Defective thymocyte maturation in p44 MAP kinase (Erk 1) knockout mice. *Science (New York, NY)* **286**: 1374-1377.

Paling NR, Wheadon H, Bone HK, Welham MJ. 2004. Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. *The Journal of biological chemistry* **279**: 48063-48070.

Park IH, Lerou PH, Zhao R, Huo H, Daley GQ. 2008. Generation of human-induced pluripotent stem cells. *Nature Protocols* **3**: 1180 - 1186.

Pera MF, Andrade J, Houssami S, Reubinoff B, Trounson A, Stanley EG, Ward-van Oostwaard D, Mummery C. 2004. Regulation of human embryonic stem cell differentiation by BMP-2 and its antagonist noggin. *Journal of cell science* **117**: 1269-1280.

Petropoulos S, Edsgärd D, Reinius B, Deng Q, Panula Sarita P, Codeluppi S, Plaza Reyes A, Linnarsson S, Sandberg R, Lanner F. 2016. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165**: 1012-1026.

Plusa B, Piliszek A, Frankenberg S, Artus J, Hadjantonakis AK. 2008. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* **135**: 3081-3091.

Pook M, Teino I, Kallas A, Maimets T, Ingerpuu S, Jaks V. 2015. Changes in Laminin Expression Pattern during Early Differentiation of Human Embryonic Stem Cells. *PloS one* **10**: e0138346.

Price PJ, Goldsborough MD, Tilkins ML. 1998. Embryonic stem cell serum replacement (KOSR patent, WO 1998030679 A1).

Rappolee DA, Basilico C, Patel Y, Werb Z. 1994. Expression and function of FGF-4 in peri-implantation development in mouse embryos. *Development* **120**: 2259-2269.

Reubinoff B, Pera MF, Fong C-Y, Trounson A, Bongso A. 2000. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nature biotechnology* **18**: 399 - 404.

Richter KS, Harris DC, Daneshmand ST, Shapiro BS. 2001. Quantitative grading of a human blastocyst: optimal ICM size and shape. *Fertility and Sterility* **76**: 1157 - 1167.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139 - 140.

Robinson MD, Smyth GK. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**: 321 - 332.

Robinson MJ, Stippec SA, Goldsmith E, White MA, Cobb MH. 1998. A constitutively active and nuclear form of the MAP kinase ERK2 is sufficient for neurite outgrowth and cell transformation. *Current biology : CB* **8**: 1141-1150.

Rodin S, Antonsson L, Hovatta O, Tryggvason K. 2014. Monolayer culturing and cloning of human pluripotent stem cells on laminin-521-based matrices under xeno-free and chemically defined conditions. *Nature Protocols* **9**: 2354-2368.

Roode M, Blair K, Snell P, Elder K, Marchant S, Smith A, Nichols J. 2012. Human hypoblast formation is not dependent on FGF signalling. *Developmental biology* **361**: 358-363.

Rossant J. 2014. Mouse and human blastocyst-derived stem cells: vive les differences. *Development* **142**: 9.

Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH. 2004. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature medicine* **10**: 55-63.

Schrode N, Saiz N, Di Talia S, Hadjantonakis AK. 2014. GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. *Developmental cell* **29**: 454-467.

Seguin CA, Draper JS, Nagy A, Rossant J. 2008. Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. *Cell stem cell* **3**: 182-195.

Shimosato D, Shiki M, Niwa H. 2007. Extra-embryonic endoderm cells derived from ES cells induced by GATA factors acquire the character of XEN cells. *BMC developmental biology* **7**: 80.

Siddle K. 2011. Signalling by insulin and IGF receptors: supporting acts and new players. *J Mol Endocrinol* **47**: R1-10.

Singh AM, Hamazaki T, Hankowski KE, Terada N. 2007. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem cells (Dayton, Ohio)* **25**: 2534-2542.

Smith A, Heath JK, Donaldson DD, Wong GG, Moreau J, Stahl M, Rogers D. 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**: 688 - 690.

Smith ER, Smedberg JL, Rula ME, Xu XX. 2004. Regulation of Ras-MAPK pathway mitogenic activity by restricting nuclear entry of activated MAPK in endoderm differentiation of embryonic carcinoma and stem cells. *The Journal of cell biology* **164**: 689-699.

Smith ZD, Chan MM, Humm KC, Karnik R, Mekhoubad S, Regev A, Eggan K, Meissner A. 2014. DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**: 611-615.

Soudais C, Bielinska M, Heikinheimo M, MacArthur CA, Narita N, Saffitz JE, Simon MC, Leiden JM, Wilson DB. 1995. Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation in vitro. *Development* **121**: 3877-3888.

Spanos S, Becker DL, Winston RML, Hardy K. 2000. Anti-Apoptotic Action of Insulin-Like Growth Factor-I During Human Preimplantation Embryo Development. *Biology of Reproduction* **63**: 1413 - 1420.

Stavridis MP, Lunn JS, Collins BJ, Storey KG. 2007. A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development* **134**: 2889-2894.

Stevens LC. 1967. Origin of testicular teratomas from primordial germ cells in mice. *Journal of the National Cancer Institute* **38**: 549 - 552.

Stevens LC, Little CC. 1954. Spontaneous Testicular Teratomas in an Inbred Strain of Mice. *Proceedings of the National Academy of Sciences of the United States of America* **40**: 1080 - 1087.

Stewart CL, Kaspar P, Brunet LJ, Bhatt H, Gadi I, Kontgen F, Abbondanzo SJ. 1992. Blastocyst implantation depends on maternal expression of LIF. *Nature* **359**: 76 - 79.

Suemori H, Yasuchika K, Hasegawa K, Fujioka T, Tsuneyoshi N, Nakatsuji N. 2006. Efficient establishment of human embryonic stem cell lines and long-term maintenance with stable karyotype by enzymatic bulk passage. *Biochemical and Biophysical Research Communications* **345**: 926 - 932.

Suzuki E, Evans T, Lowry J, Truong L, Bell DW, Testa JR, Walsh K. 1996. The Human GATA-6 Gene: Structure, Chromosomal Location, and Regulation of Expression by Tissue-Specific and Mitogen-Responsive Signals. *Genomics* **38**: 283 - 290.

Takahashi K, Mitsui K, Yamanaka S. 2003. Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. *Nature* **423**: 541-545.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**: 663-676.

Takashima Y, Guo G, Loos R, Nichols J, Ficz G, Krueger F, Oxley D, Santos F, Clarke J, Mansfield W et al. 2014. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**: 1254-1269.

Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. 2011. Differential expression in RNA-seq: a matter of depth. *Genome Research* **21**: 2213 - 2223.

Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD. 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**: 196-199.

Theunissen TW, Friedli M, He Y, Planet E, O'Neill RC, Markoulaki S, Pontis J, Wang H, Iouranova A, Imbeault M et al. 2016. Molecular criteria for defining the naive human pluripotent state. *Cell stem cell* **19**: 502 - 515.

Theunissen TW, Powell BE, Wang H, Mitalipova M, Faddah DA, Reddy J, Fan ZP, Maetzel D, Ganz K, Shi L et al. 2014. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell stem cell* **15**: 471-487.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. 1998. Embryonic stem cell lines derived from human blastocysts. *Science (New York, NY)* **282**: 1145-1147.

Thomson JA, Kalishman J, Golos TG, Durning M, Harris CP, Becker RA, Hearn JP. 1995. Isolation of a primate embryonic stem cell line. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 7844 - 7848.

Thomson JA, Kalishman J, Golos TG, Durning M, Harris CP, Hearn JP. 1996. Pluripotent Cell Lines Derived from Common Marmoset (Callithrix jacchus) Blastocysts. *Biology of Reproduction* **55**: 254 - 259.

Tiyaboonchai A, Cardenas-Diaz FL, Ying L, Maguire JA, Sim X, Jobaliya C, Gagne AL, Kishore S, Stanescu DE, Hughes N et al. 2017. GATA6 Plays an Important Role in the Induction of Human Definitive Endoderm, Development of the Pancreas, and Functionality of Pancreatic beta Cells. *Stem cell reports* **8**: 1 - 16.

Traverse S, Gomez N, Paterson H, Marshall C, Cohen P. 1992. Sustained activation of the mitogen-activated protein (MAP) kinase cascade may be required for differentiation of PC12 cells. Comparison of the effects of nerve growth factor and epidermal growth factor. *The Biochemical journal* **288 (Pt 2)**: 351-355.

Uranishi K, Akagi T, Sun C, Koide H, Yokota T. 2013. Dax1 Associates with Esrrb and Regulates Its Function in Embryonic Stem Cells. *Molecular and cellular biology* **33**: 2056 - 2066.

Valamehr B, Robinson M, Abujarour R, Rezner B, Vranceanu F, Le T, Medcalf A, Lee TT, Fitch M, Robbins D et al. 2014. Platform for induction and maintenance of transgene-free hiPSCs resembling ground state pluripotent stem cells. *Stem cell reports* **2**: 366-381.

Valera E, Isaacs MJ, Kawakami Y, Izpisua Belmonte JC, Choe S. 2010. BMP-2/6 heterodimer is more effective than BMP-2 or BMP-6 homodimers as inductor of differentiation of human embryonic stem cells. *PloS one* **5**: e11167.

Vallier L, Alexander M, Pedersen RA. 2005. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *Journal of cell science* **118**: 4495-4509.

Vallier L, Reynolds D, Pedersen RA. 2004. Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Developmental biology* **275**: 403 - 421.

Vallier L, Touboul T, Brown S, Cho C, Bilican B, Alexander M, Cedervall J, Chandran S, Ahrlund-Richter L, Weber A et al. 2009. Signaling pathways controlling pluripotency and early cell fate decisions of human induced pluripotent stem cells. *Stem cells (Dayton, Ohio)* **27**: 2655-2666.

van den Berg DL, Zhang W, Yates A, Engelen E, Takacs K, Bezstarosti K, Demmers J, Chambers I, Poot RA. 2008. Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. *Molecular and cellular biology* **28**: 5986-5995.

Van der Jeught M, Heindryckx B, O'Leary T, Duggal G, Ghimire S, Lierman S, van Roy F, Chuva de Sousa Lopes SM, Deroo T, Deforce D et al. 2013. Treatment of human embryos with the TGFb inhibitor SB431542 increases epiblast proliferation and permits successful human embryonic stem cell derivation. *Human Reproduction* **29**: 41 - 48.

van Vliet J, Crofts LA, Quinlan KG, Czolij R, Perkins AC, Crossley M. 2006. Human KLF17 is a new member of the Sp/KLF family of transcription factors. *Genomics* **87**: 474-482.

Vokes SA, Ji H, McCuine S, Tenzen T, Giles S, Zhong S, Longabaugh WJ, Davidson EH, Wong WH, McMahon AP. 2007. Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. *Development* **134**: 1977-1989.

Wamaitha SE, Del Valle I, Cho LT, Wei Y, Fogarty NM, Blakeley P, Sherwood RI, Ji H, Niakan KK. 2015. Gata6 potently initiates reprograming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. *Genes & development* **29**: 1239 - 1255.

Wang L, Schulz TC, Sherrer ES, Dauphin DS, Shin S, Nelson AM, Ware CB, Zhan M, Song C-Z, Chen X et al. 2007. Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling. *Blood* **110**: 4111 - 4119.

Wang Y, Smedberg JL, Cai KQ, Capo-Chichi DC, Xu XX. 2011. Ectopic expression of GATA6 bypasses requirement for Grb2 in primitive endoderm formation. *Developmental dynamics : an official publication of the American Association of Anatomists* **240**: 566-576.

Wang Z, Oron E, Nelson B, Razis S, Ivanova N. 2012. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell stem cell* **10**: 440-454.

Ware CB, Nelson AM, Mecham B, Hesson J, Zhou W, Jonlin EC, Jimenez-Caliani AJ, Deng X, Cavanaugh C, Cook S et al. 2014. Derivation of naive human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 4484-4489.

Watanabe S, Umehara H, Murayama K, Okabe M, Kimura T, Nakano T. 2006. Activation of Akt signaling is sufficient to maintain pluripotency in mouse and primate embryonic stem cells. *Oncogene* **25**: 2697-2707.

Weinberger L, Ayyash M, Novershtern N, Hanna JH. 2016. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat Rev Mol Cell Biol* **17**: 155-169.

Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, Wagner EF, Metcalf D, Nicola NA, Gough NM. 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**: 684 - 687.

Wray J, Kalkan T, Gomez-Lopez S, Eckardt D, Cook A, Kemler R, Smith A. 2011. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nature Cell Biology* **13**: 838-845.

Xin M, Davis CA, Molkentin JD, Lien CL, Duncan SA, Richardson JA, Olson EN. 2006. A threshold of GATA4 and GATA6 expression is required for cardiovascular development. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 11189 - 11194.

Xu C, Inokuma MS, Denham J, Golds K, Kundu P, Gold JD, Carpenter MK. 2001. Feeder-free growth of undifferentiated human embryonic stem cells. *Nature biotechnology* **19**: 971 - 974.

Xu RH, Chen X, Li DS, Li R, Addicks GC, Glennon C, Zwaka TP, Thomson JA. 2002. BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nature biotechnology* **20**: 1261-1264.

Xu RH, Peck RM, Li DS, Feng X, Ludwig TE, Thomson JA. 2005. Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nature Methods* **2**: 185 - 190.

Yamanaka Y, Lanner F, Rossant J. 2010. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**: 715-724.

Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J et al. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature structural & molecular biology* **20**: 1131-1139.

Yayon A, Klagsbrun M, Esko JD, Leder P, Ornitz DM. 1991. Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor. *Cell* **64**: 841-848.

Yeo JC, Jiang J, Tan ZY, Yim GR, Ng JH, Goke J, Kraus P, Liang H, Gonzales KA, Chong HC et al. 2014. Klf2 is an essential factor that sustains ground state pluripotency. *Cell stem cell* **14**: 864-872.

Yeom YI, Fuhrmann G, Ovitt CE, Brehm A, Ohbo K, Gross M, Hubner K, Scholer HR. 1996. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**: 881.

Ying QL, Smith A. 2012. Culture medium containing kinase inhibitors, and uses thereof (2i patent, EP 2457995 A1).

Ying QL, Stavridis M, Griffiths D, Li M, Smith A. 2003. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature biotechnology* **21**: 183-186.

Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A. 2008. The ground state of embryonic stem cell self-renewal. *Nature* **453**: 519-523.

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R et al. 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science (New York, NY)* **318**: 1917 - 1920.

Yu M, Riva L, Xie H, Schindler Y, Moran TB, Cheng Y, Yu D, Hardison R, Weiss MJ, Orkin SH et al. 2009. Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Molecular Cell* **36**: 682 - 695.

Zaret KS, Carroll JS. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes & development* **25**: 2227 - 2241.

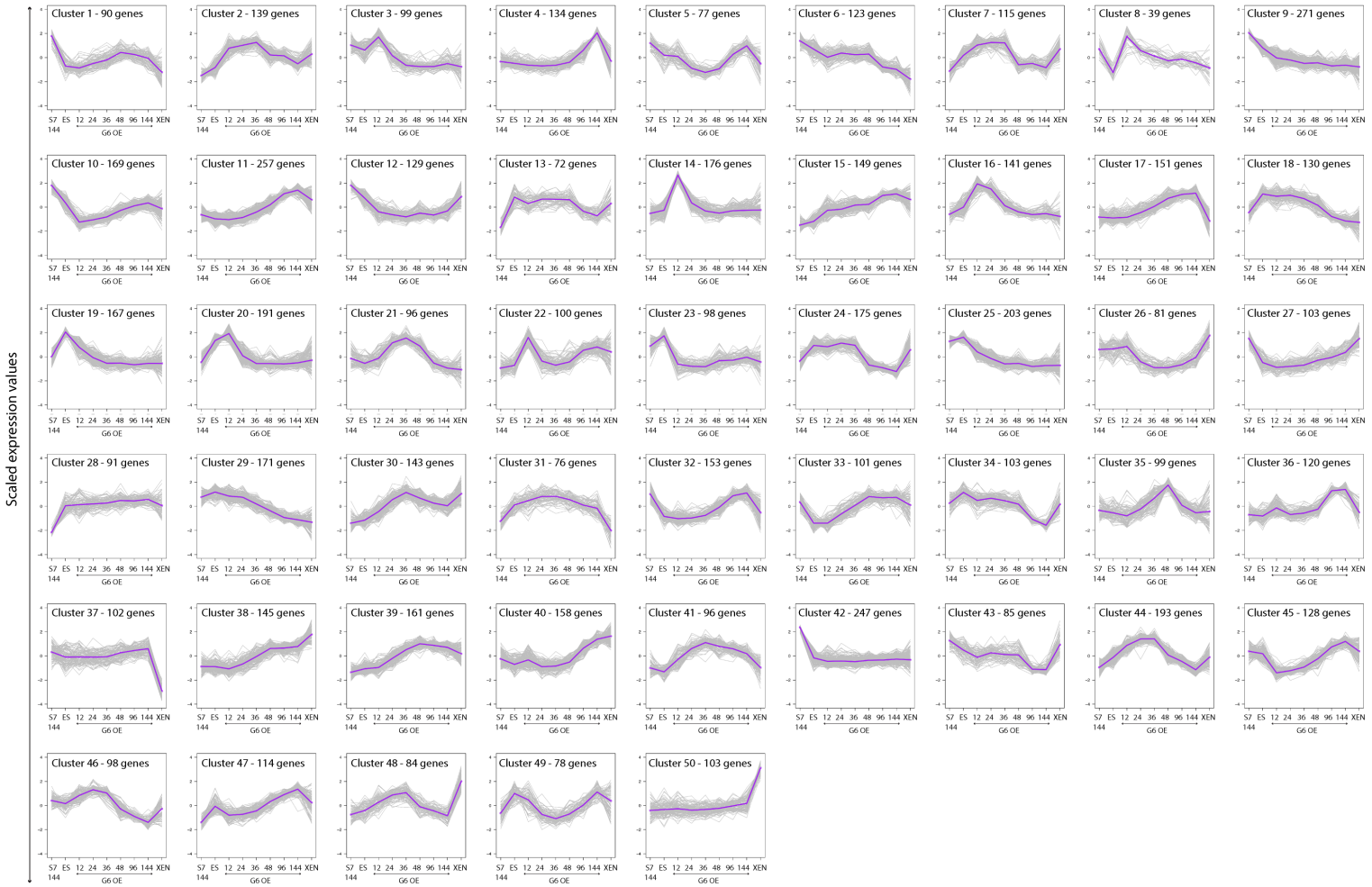
Zhang P, Andrianakos R, Yang Y, Liu C, Lu W. 2010. Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. *The Journal of biological chemistry* **285**: 9180-9189.

Zhang X, Zhang J, Wang T, Esteban MA, Pei D. 2008. Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *The Journal of biological chemistry* **283**: 35825-35833.

Zhao M, Ren C, Yang H, Feng X, Jiang X, Zhu B, Zhou W, Wang L, Zeng Y, Yao K. 2007. Transcriptional profiling of human embryonic stem cells and embryoid bodies identifies HESRG, a novel stem cell gene. *Biochemical and Biophysical Research Communications* **362**: 916 - 922.

Zhou Q, Chipperfield H, Melton DA, Wong WH. 2007. A gene regulatory network in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 16438 - 16443.

Appendix



Appendix Figure A1. Microarray gene expression clusters. Line plots generated following microarray analysis of *Gata6*-induced cells from 12 to 144 hours following doxycycline treatment. Uninduced mES cells, embryo-derived XEN cells and *Sox7*-induced cells after 144 hours of doxycycline treatment were also included. Genes were grouped into 50 clusters using K-means clustering, according to normalized gene expression values scaled across time points. The trajectory of scaled expression values for individual genes in each cluster across time are shown as grey lines, while the purple lines correspond to the cluster mean.

GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b
GO:0060071	Wnt signaling pathway, planar cell polarity pathway	7.27E-11	8.28E-09	3.06	2946	46	690	33
GO:0035567	non-canonical Wnt signaling pathway	3.16E-10	3.38E-08	2.87	2946	52	690	35
GO:0002764	immune response-regulating signaling pathway	1.41E-09	1.42E-07	2.67	2946	78	581	41
GO:0002757	immune response-activating signal transduction	3.54E-09	3.44E-07	2.67	2946	74	581	39
GO:0038093	Fc receptor signaling pathway	5.17E-09	4.93E-07	2.85	2946	67	540	35
GO:0033209	tumor necrosis factor-mediated signaling pathway	9.27E-09	8.65E-07	2.87	2946	41	725	29
GO:0090263	positive regulation of canonical Wnt signaling pathway	2.26E-08	2.04E-06	3.11	2946	44	581	27
GO:0002768	immune response-regulating cell surface receptor signaling pathway	2.29E-08	2.05E-06	2.69	2946	73	540	36
GO:0002758	innate immune response-activating signal transduction	3.74E-08	3.25E-06	2.94	2946	50	581	29
GO:0002429	immune response-activating cell surface receptor signaling pathway	6.65E-08	5.55E-06	2.69	2946	69	540	34
GO:0038061	NIK/NF-kappaB signaling	1.07E-07	8.68E-06	3.09	2946	41	581	25
GO:0038095	Fc-epsilon receptor signaling pathway	1.07E-07	8.63E-06	2.9	2946	49	581	28
GO:0030177	positive regulation of Wnt signaling pathway	3.12E-07	2.28E-05	2.58	2946	49	699	30
GO:0050852	T cell receptor signaling pathway	4.99E-07	3.45E-05	2.87	2946	46	581	26
GO:0019221	cytokine-mediated signaling pathway	5.05E-07	3.46E-05	2.32	2946	63	725	36
GO:0002223	stimulatory C-type lectin receptor signaling pathway	1.46E-06	9.02E-05	2.82	2946	45	581	25
GO:0002220	innate immune response activating cell surface receptor signaling pathway	1.46E-06	8.95E-05	2.82	2946	45	581	25
GO:0050851	antigen receptor-mediated signaling pathway	1.58E-06	9.58E-05	2.75	2946	48	581	26
GO:0090090	negative regulation of canonical Wnt signaling pathway	1.89E-06	1.12E-04	2.47	2946	49	705	29
GO:0030178	negative regulation of Wnt signaling pathway	4.38E-06	2.40E-04	2.37	2946	53	705	30
GO:0060828	regulation of canonical Wnt signaling pathway	6.85E-06	3.60E-04	2.27	2946	59	705	32
GO:0000165	MAPK cascade	2.11E-05	9.57E-04	2.42	2946	63	540	28
GO:0030111	regulation of Wnt signaling pathway	3.83E-05	1.65E-03	2.06	2946	71	705	35
GO:0016055	Wnt signaling pathway	7.82E-05	3.09E-03	2.11	2946	88	540	34
GO:1905114	cell surface receptor signaling pathway involved in cell-cell signaling	7.82E-05	3.08E-03	2.11	2946	88	540	34
GO:0017015	regulation of transforming growth factor beta receptor signaling pathway	8.14E-05	3.19E-03	3.87	2946	16	524	11
GO:0023014	signal transduction by protein phosphorylation	1.18E-04	4.35E-03	2.25	2946	68	540	28
GO:1903844	regulation of cellular response to transforming growth factor beta stimulus	1.89E-04	6.57E-03	3.64	2946	17	524	11
GO:0030512	negative regulation of transforming growth factor beta receptor signaling pathway	1.97E-04	6.79E-03	4.5	2946	10	524	8
GO:0002755	MyD88-dependent toll-like receptor signaling pathway	2.69E-04	8.91E-03	6.65	2946	5	443	5
GO:0090092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	5.79E-04	1.66E-02	2.83	2946	25	583	14
GO:1903845	negative regulation of cellular response to transforming growth factor beta stimulus	6.50E-04	1.82E-02	4.09	2946	11	524	8

Appendix Table X1. Selected gene ontology analysis categories enriched in the Epi. Gene ontology analysis of all the genes expressed above 5 RPKM in half or more Epi cells.

GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b
GO:0060071	Wnt signaling pathway, planar cell polarity pathway	1.10E-18	1.94E-16	4.24	3732	46	727	38
GO:0035567	non-canonical Wnt signaling pathway	1.68E-16	2.57E-14	3.8	3732	54	727	40
GO:0033209	tumor necrosis factor-mediated signaling pathway	1.35E-13	1.77E-11	3.61	3732	46	764	34
GO:0038061	NIK/NF-kappaB signaling	4.33E-12	4.81E-10	3.62	3732	44	727	31
GO:0090263	positive regulation of canonical Wnt signaling pathway	6.64E-11	6.59E-09	3.39	3732	47	727	31
GO:0002764	immune response-regulating signaling pathway	7.88E-11	7.66E-09	2.76	3732	109	582	47
GO:0002758	innate immune response-activating signal transduction	9.75E-11	9.38E-09	3.35	3732	65	582	34
GO:0002757	immune response-activating signal transduction	2.64E-10	2.35E-08	2.79	3732	101	582	44
GO:0002768	immune response-regulating cell surface receptor signaling pathway	3.99E-10	3.45E-08	2.91	3732	97	542	41
GO:0038093	Fc receptor signaling pathway	4.22E-10	3.58E-08	3.04	3732	86	542	38
GO:0002223	stimulatory C-type lectin receptor signaling pathway	4.69E-10	3.91E-08	3.07	3732	53	758	33
GO:0002220	innate immune response activating cell surface receptor signaling pathway	4.69E-10	3.88E-08	3.07	3732	53	758	33
GO:0038095	Fc-epsilon receptor signaling pathway	5.75E-10	4.55E-08	3.44	3732	62	542	31
GO:0050851	antigen receptor-mediated signaling pathway	6.86E-10	5.17E-08	3.5	3732	59	542	30
GO:0090090	negative regulation of canonical Wnt signaling pathway	7.77E-10	5.80E-08	3.51	3732	53	582	29
GO:0002429	immune response-activating cell surface receptor signaling pathway	1.46E-09	1.06E-07	2.94	3732	89	542	38
GO:0030178	negative regulation of Wnt signaling pathway	1.47E-09	1.06E-07	2.96	3732	59	727	34
GO:0030177	positive regulation of Wnt signaling pathway	2.55E-09	1.80E-07	3.38	3732	55	582	29
GO:0016055	Wnt signaling pathway	2.45E-08	1.50E-06	2.63	3732	103	551	40
GO:1905114	cell surface receptor signaling pathway involved in cell-cell signaling	2.45E-08	1.49E-06	2.63	3732	103	551	40
GO:0019221	cytokine-mediated signaling pathway	4.48E-08	2.57E-06	2.59	3732	68	764	36
GO:0060828	regulation of canonical Wnt signaling pathway	1.08E-07	5.92E-06	3.03	3732	66	542	29
GO:0000165	MAPK cascade	2.24E-06	1.04E-04	2.43	3732	82	637	34
GO:0030111	regulation of Wnt signaling pathway	3.00E-06	1.36E-04	2.57	3732	83	542	31
GO:0002755	MyD88-dependent toll-like receptor signaling pathway	1.50E-04	4.80E-03	9.2	3732	6	338	5
GO:0009967	positive regulation of signal transduction	3.88E-04	1.10E-02	1.6	3732	280	542	65
GO:0017015	regulation of transforming growth factor beta receptor signaling pathway	4.08E-04	1.15E-02	3.94	3732	19	499	10
GO:1903844	regulation of cellular response to transforming growth factor beta stimulus	4.08E-04	1.14E-02	3.94	3732	19	499	10
GO:0009968	negative regulation of signal transduction	4.24E-04	1.18E-02	1.63	3732	254	542	60
GO:0030512	negative regulation of transforming growth factor beta receptor signaling pathway	4.63E-04	1.27E-02	5.28	3732	11	450	7
GO:1903845	negative regulation of cellular response to transforming growth factor beta stimulus	4.63E-04	1.27E-02	5.28	3732	11	450	7
GO:0023056	positive regulation of signaling	5.63E-04	1.50E-02	1.56	3732	300	542	68

Appendix Table X2. Selected gene ontology analysis categories enriched in the TE. Gene ontology analysis of all the genes expressed above 5 RPKM in half or more TE cells.